Accrediting outputs of noisy intermediate-scale quantum computing devices

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG IOP Institute of Physics

PAPER • OPEN ACCESS

Accrediting outputs of noisy intermediate-scale quantum computing devices

To cite this article: Samuele Ferracin *et al* 2019 *New J. Phys.* **21** 113038

IOP Publishing New J. Phys. 21 (2019) 113038 https://doi.org/10.1088/1367-2630/ab4fd6

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

CrossMark

OPEN ACCESS

RECEIVED
28 June 2019

REVISED
3 October 2019

ACCEPTED FOR PUBLICATION
21 October 2019

PUBLISHED
18 November 2019

PAPER

Accrediting outputs of noisy intermediate-scale quantum computing devices

Samuele Ferracin, Theodoros Kapourniotis and Animesh Datta

Department of Physics, University of Warwick, Coventry CV4 7AL, United Kingdom

E-mail: animesh.datta@warwick.ac.uk, s.ferracin@warwick.ac.uk and t.kapourniotis@warwick.ac.uk

Keywords: NISQ devices, accreditation, sampling problems

Abstract

We present an accreditation protocol for the outputs of noisy intermediate-scale quantum devices. By testing entire circuits rather than individual gates, our accreditation protocol can provide an upper-bound on the variation distance between noisy and noiseless probability distribution of the outputs of the target circuit of interest. Our accreditation protocol requires implementing quantum circuits no larger than the target circuit, therefore it is practical in the near term and scalable in the long term. Inspired by trap-based protocols for the verification of quantum computations, our accreditation protocol assumes that single-qubit gates have bounded probability of error. We allow for arbitrary spatial and temporal correlations in the noise affecting state preparation, measurements, single-qubit and two-qubit gates. We describe how to implement our protocol on real-world devices, and we also present a novel cryptographic protocol (which we call 'mesothetic' protocol) inspired by our accreditation protocol.

1. Introduction

Quantum computers promise to expand our computing capabilities beyond their current horizons. Several commercial institutions [1–3] are taking steps towards building the first prototypes of quantum computers that can outperform existing supercomputers in certain tasks [4–9], the so-called 'Noisy Intermediate-Scale Quantum' (NISQ) computing devices [10]. As all their internal operations such as state preparations, gates, and measurements are by definition noisy, the outputs of computations implemented on NISQ devices are unreliable. It is thus essential to devise protocols able to accredit these outputs.

A commonly employed approach involves simulating the quantum circuit whose output we wish to accredit, the target circuit, on a classical computer. This is feasible for small circuits, as well as for circuits composed of Clifford gates [11] and few non-Clifford gates [12, 13]. Classical simulations have been performed for quantum computations of up to 72 qubits, often exploiting subtle insights into the nature of specific quantum circuits involved [14, 15]. Though practical for the present, classical simulations of quantum circuits are not scalable. Worthwhile quantum computations will not be efficiently simulable on classical computers, hence we must seek for alternative methods.

Another approach employed in experiments consists of individually testing classes of gates present in the target circuit. This is typically undertaken using a family of protocols centered around randomized benchmarking and its extensions [16–21]. These protocols allow extraction of the fidelity of gates or cycles of gates and can witness progresses towards fault-tolerant quantum computing [22]. However they rely on assumptions that may be invalid in experiments. In particular, they require the noise to be Markovian and cannot account for temporal correlations [23, 24]. Quantum circuits are more than the sum of their gates, and the noise in the target circuit may exhibit characteristics that cannot be captured by benchmarking its individual gates independently.

This calls for protocols able to test circuits as a whole rather than individual gates. Such protocols have been devised inspired by Interactive proof Systems [25]. In these protocols (which we call 'cryptographic protocols') the outputs of the target circuit are verified through an interaction between a trusted verifier and an untrusted prover (figure 1(a)). The verifier is typically allowed to possess a noiseless quantum device able to prepare [26–33] or

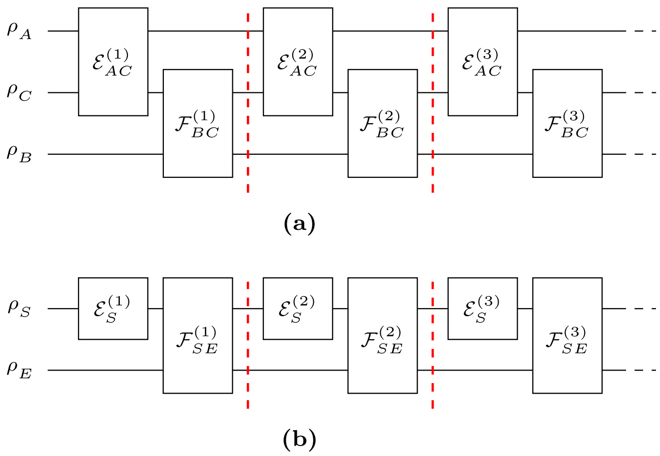

Figure 1. (a) In cryptographic protocols a verifier A and a prover B apply operations on their own registers and on a shared register C. (b) In accreditation protocols all operations applied to the system S are noisy. Noise couples the system to an environment E.

measure [34–38] single qubits, however recently a protocol for a fully classical verifier was devised that relies on the widely believed intractability of a computational problem for quantum computers [39]. Other protocols for classical verifiers have also been devised, but they require interaction with multiple entangled and non-communicating provers [40–44]. Cryptographic protocols show that with minimal assumptions, verification of the outputs of quantum computations of arbitrary size can be done efficiently, in principle.

In practice, implementing cryptographic protocols in experiments remains challenging, especially in the near term. In experiments all the operations are noisy, as in figure 1(b), and the verifier does not possess noiseless quantum devices. Thus, the verifiability of protocols requiring noiseless devices for the verifier is not guaranteed. Moreover, the concept of *scalability*, which is of primary interest in cryptographic protocols, is not equivalent to that of *practicality*, which is essential for experiments. For instance, suppose that the target circuit contains a few hundred qubits and a few hundred gates. Cryptographic protocols require implementing this circuit on a large cluster state containing thousands of qubits and entangling gates [26–31, 34] or on two spatially-separated devices sharing thousands of copies of Bell states [40, 41]; or appending several teleportation gadgets to the target circuit (one for each T-gate in the circuit and six for each Hadamard gate) [33]; or building Feynman–Kitaev clock states, which require entangling the system with an auxiliary qubit per gate in the target circuit [35, 39, 43]. These protocols are scalable, as they require a number of additional qubits, gates and measurements growing linearly with the size of the target circuit, yet they remain impractical for NISQ devices.

In this paper we present an accreditation protocol that provides an upper-bound on the variation distance between noisy and noiseless probability distribution of the outputs of a NISQ device, under the assumption that

N1: Noise in state preparation, entangling gates, and measurements is an arbitrary Completely Positive Trace Preserving (CPTP) map encompassing the whole system and the environment (equation (2));

N2: Noise in single-qubit gates is a CPTP map $\mathcal{F}_{SE}$ of the form $\mathcal{F}_{SE} = (1-r)\mathcal{I} + r\mathcal{F}'_{SE}$ with $0 \leqslant r < 1$, where $\mathcal{I}_{SE}$ is the identity on system and environment and $\mathcal{F}'_{SE}$ is an arbitrary (potentially gate-dependent) CPTP map encompassing the whole system and the environment.

Inspired by cryptographic protocols [26–44], our accreditation protocol is trap-based, meaning that the target circuit being accredited is implemented together with a number v of classically simulable circuits (the 'trap' circuits) able to detect all types of noise subject to conditions N1 and N2 above. A single run of our protocol requires implementing the target circuit being accredited and v trap circuits. It provides a binary outcome in which the outputs of the target circuit are either accepted as *correct* (with confidence increasing linearly with v) or rejected as *potentially incorrect*. More usefully, consider running our protocol d times, each time with the same target and v potentially different trap circuits. Suppose that the output of the target is accepted as correct by $N_{\rm acc} > 0$ runs. With confidence $1 - 2\exp(-2d\theta^2)$, for each of these accepted outputs our protocol ensures

$$\frac{1}{2}\sum_{\bar{s}}|p_{\rm noiseless}(\bar{s}) - p_{\rm noisy}(\bar{s})| \leqslant \frac{\varepsilon}{N_{\rm acc}/d - \theta}, \tag{1}$$

where $\theta \in (0, N_{\rm acc}/d)$ is a tunable parameter that affects both the confidence and the upper-bound, $p_{\rm noiseless}(\bar{s})$ and $p_{\rm noisy}(\bar{s})$ are the noiseless and noisy probability distributions of the outputs $\{\bar{s}\}$ of the target circuit respectively and $\varepsilon \propto 1/v$. Bounds of this type can fruitfully accredit the outputs of experimental quantum computers as well as underpin attempts at demonstrating and verifying quantum supremacy in sampling experiments [4–9].

Crucially, our accreditation protocol is both experimentally practical and scalable: all circuits implemented in our protocol are no wider (in the number of qubits) or deeper (in the number of gates) than the circuit we seek to accredit. This makes our protocol more readily implementable on NISQ devices than cryptographic protocols. Moreover, our protocol requires no noiseless quantum device, and it only relies on the assumption that the single-qubit gates suffer bounded (but potentially non-local in space and time and gate-dependent) noise—condition N2. This assumption is motivated by the empirical observation that single-qubit gates are the most accurate operations in prominent quantum computing platforms such as trapped ions [45, 46] and superconducting qubits [2, 47, 48].

In addition to its ready implementability on NISQ devices, our accreditation protocol can detect all types of noise typically considered by techniques centered around randomized benchmarking and its extensions [16–21]. Moreover it can detect noise that may be missed by those techniques such as noise correlated in time. Mathematically, this amounts to allowing noisy operations to encompass both system and environment (figure 1(b)) and tracing out the environment only at the end of the protocol. This noise model is more general than the Markovian noise model considered in protocols centered around randomized benchmarking [23, 24]. Moreover, by testing circuits rather than gates, our protocol ensures that all possible noise (subject to condition N1 and N2) in state preparation, measurement and gates is detected, even noise that arises only when these components are put together to form a circuit. On the contrary, benchmarking isolated gates can sometimes yield over-estimates of their fidelities [21], and consequently of the fidelity of the resulting circuit. We note that noise of the type N2 excludes unbounded gate-dependent errors in single-qubit gates such as systematic over- or under-rotations, as also is the case for other works [16–21].

Inspired by our accreditation protocol we also present a novel cryptographic protocol, which we call 'mesothetic verification protocol'. In the mesothetic protocol the verifier implements the single-qubit gates in all circuits while the prover undertakes all other operations. This is distinct from prepare-and-send [26–33] or receive-and-measure [34–38] cryptographic protocols in that the verifier intervenes during the actual implementation of the circuits, and not before or after the circuits are implemented.

Our paper is organized as follows. In section 1 of Results we introduce the notation, in section 2 we provide the necessary definitions, in sections 3 and 4 we present our protocol and prove our results and in section 2.5 we present the mesothetic verification protocol.

2. Results

2.1. Notation

We indicate unitary matrices acting on the system with capital letters such as U, V and W, and Completely Positive Trace Preserving (CPTP) maps with calligraphic letters such as $\mathcal{E}$, $\mathcal{F}$, $\mathcal{R}$ and $\mathcal{M}$. We indicate the 2×2 identity matrix as I, the single-qubit Pauli gates as X, Y, Z, the controlled-Z gate as cZ, the controlled-X gate as cX, the Hadamard gate as H and $S = \mathrm{diag}(1, i)$. The symbol $\circ$ denotes the composition of CPTP maps: $\circ_{p=1}^{q}\mathcal{E}_p(\rho) = \mathcal{E}_q \ldots \mathcal{E}_1(\rho)$, $\mathrm{Tr}_E[\cdot]$ is the trace over the environment, $D(\sigma, \tau) = \mathrm{Tr}|\sigma - \tau|/2$ is the trace distance between the states σ and τ. We say that a noisy implementation $\widetilde{\mathcal{E}}$ of $\mathcal{E}$ suffers bounded noise if $\widetilde{\mathcal{E}}$ can be written as $\widetilde{\mathcal{E}} = (1 - r)\mathcal{E} + r\mathcal{F}$ for some CPTP map $\mathcal{F}$ and number $0 \leqslant r < 1$, otherwise if $r = 1$ we say that the noise is unbounded [49].

2.2. Background

We start by defining our notion of protocol:

Definition 1 (Protocol). Consider a system S in the state ρ_S. A protocol on input ρ_S is a collection of CPTP maps $\{\mathcal{E}_S^{(p)}\}_{p=1}^{q}$ acting on S and yielding the state $\rho_{\mathrm{out}} = \circ_{p=1}^{q}\mathcal{E}_S^{(p)}(\rho_S)$. When implemented on real devices protocols suffer the effects of noise. Modeling noise as a set $\{\mathcal{F}_{SE}^{(p)}\}$ of CPTP maps acting on system and environment (figure 1(b)), the state of the system at the end of a noisy protocol run is

$$\rho_{\mathrm{out}} = \mathrm{Tr}_E[\circ_{p=1}^{q}\mathcal{F}_{SE}^{(p)}(\mathcal{E}_S^{(p)} \otimes \mathcal{I}_E)(\rho_S \otimes \rho_E)], \tag{2}$$

where ρ_E is the state of the environment at the beginning of the protocol. We allow each map $\mathcal{F}_{SE}^{(p)}$ to depend arbitrarily on the corresponding operation $\mathcal{E}_S^{(p)}$.

A trap-based accreditation protocol is defined as follows. A single run of such a protocol takes as input a classical description of the target circuit and a number v, implements $v + 1$ circuits (the target and v traps) and returns the outputs of the target circuit, together with a 'flag bit' set to 'acc' ('rej') indicating that the output of the target must be accepted (rejected). Formally,

Definition 2(Trap-Based Accreditation Protocol). Consider a protocol $\{\mathcal{E}_S^{(p)}\}_{p=1}^q$ with input ρ_S, where ρ_S contains a classical description of the target circuit and the number v of trap circuits. Consider also a set of CPTP maps $\{\mathcal{F}_{SE}^{(p)}\}_{p=1}^q$ (the noise) acting on system and environment. We say that the protocol $\{\mathcal{E}_S^{(p)}\}_{p=1}^q$ can accredit the outputs of the target circuit in the presence of noise $\{\mathcal{F}_{SE}^{(p)}\}_{p=1}^q$ if the following two properties hold:

(1) The state of the system at the end of a single protocol run (equation (2)) can be expressed as

$$\rho_{\mathrm{out}} = b\ \tau_{\mathrm{out}}^{\prime\mathrm{tar}} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}| + (1-b)(l\ \sigma_{\mathrm{out}}^{\mathrm{tar}} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}| + (1-l)\tau_{\mathrm{out}}^{\mathrm{tar}} \otimes |\mathrm{rej}\rangle\langle\mathrm{rej}|), \quad (3)$$

where $\sigma_{\mathrm{out}}^{\mathrm{tar}}$ $(\tau_{\mathrm{out}}^{\prime\mathrm{tar}})$ is the state of the target circuit at the end of a noiseless (noisy) protocol run, $\tau_{\mathrm{out}}^{\mathrm{tar}}$ is an arbitrary state for the target circuit, $|\mathrm{acc}\rangle$ is the state of the flag indicating acceptance, $|\mathrm{rej}\rangle = |\mathrm{acc} \oplus 1\rangle$, $0 \leqslant l \leqslant 1, 0 \leqslant b \leqslant \varepsilon$ and $\varepsilon \in [0,\ 1]$.

(2) After d protocol runs with the same target circuit and v potentially different trap circuits, if all these runs are affected by independent and identically distributed (i.i.d.) noise, then the variation distance between noisy and noiseless probability distribution of the outputs of each of the $N_{\mathrm{acc}} \in [0,\ d]$ protocol runs ending with flag bit in the state $|\mathrm{acc}\rangle$ is upper-bounded as in equation (1).

Property 1 ensures that the probability of accepting the outputs of a single protocol run when the target circuit is affected by noise (the number b in equation (3)) is smaller than a constant ε. The constant ε is a function of the number of trap circuits, of the protocol and of the noise model and is to be computed analytically. The quantity $1-\varepsilon$ quantifies the *credibility* of the accreditation protocol.

Note that Property 1 in the above definition implies Property 2. To see this, assume Property 1 is valid for a given protocol. Suppose that this protocol is run d times with i.i.d. noise (a standard assumption in trap-based cryptographic protocols [29, 36]) and suppose that $N_{\mathrm{acc}} > 0$ protocol runs end with flag bit in the state $|\mathrm{acc}\rangle$. For each of these N_{acc} runs, the state of the system at the end of the protocol run is thus of the form (see equation (3))

$$\rho_{\mathrm{out,acc}} = \frac{(1-b)l\ \sigma_{\mathrm{out}}^{\mathrm{tar}} + b\ \tau_{\mathrm{out}}^{\prime\mathrm{tar}}}{(1-b)l + b} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}|. \quad (4)$$

This yields a bound on the variation distance of the type [50]

$$\begin{aligned} &\frac{1}{2}\sum_{\bar{s}}|p_{\mathrm{noiseless}}(\bar{s}) - p_{\mathrm{noisy}}(\bar{s})| \\ &\leqslant D\left(\sigma_{\mathrm{out}}^{\mathrm{tar}}, \frac{(1-b)l\ \sigma_{\mathrm{out}}^{\mathrm{tar}} + b\ \tau_{\mathrm{out}}^{\prime\mathrm{tar}}}{(1-b)l + b}\right) \\ &\leqslant \frac{b}{(1-b)l + b} \leqslant \frac{\varepsilon}{\mathrm{prob(acc)}}, \end{aligned} \quad (5)$$

where in the last inequality we used that $b \leqslant \varepsilon$ (Property 1) and that the quantity prob(acc)=$(1-b)l + b$ is the probability of accepting (equation (3)). Hoeffding's Inequality ensures that $|\mathrm{prob(acc)} - N_{\mathrm{acc}}/d| \leqslant \theta$ with confidence $1 - 2\exp(-2d\theta^2)$ and this yields Property 2.

Bounding the variation distance as in equation (1) requires knowledge of the two numbers ε and N_{acc}, the former obtained theoretically from the protocol and the latter experimentally from the device being tested. ε is a property of the protocol, of its input and of the noise model and can be computed without running the protocol. However, different devices running the same target circuit will suffer different noise levels and this is captured by N_{acc}, which depends on the experimental device being tested. It is important to note that the bound on the variation distance is valid only for the outputs of the N_{acc} protocol runs ending with flag bit in the state $|\mathrm{acc}\rangle$. If a protocol run ends with flag bit in the state $|\mathrm{rej}\rangle$, Property 1 implies no bound on the variation distance and all rejected outputs must be discarded.

We can now present our accreditation protocol (a formal description can be found in Box 1 in the Methods).

2.3. Our accreditation protocol

Our accreditation protocol takes as input a classical description of the target circuit and the number v of trap circuits. The target circuit (figure 2) must start with qubits in the state $|+\rangle$, contain only single-qubit gates and cZ gates and end with a round of measurements in the Pauli-X basis[1]. Moreover, it must be decomposed as a

[1] This does not result in any loss of generality: every experimental architecture has its native input states, entangling gates and measurement basis, but these can always be mapped to $|+\rangle$ states, cZ gates and Pauli-X measurements.

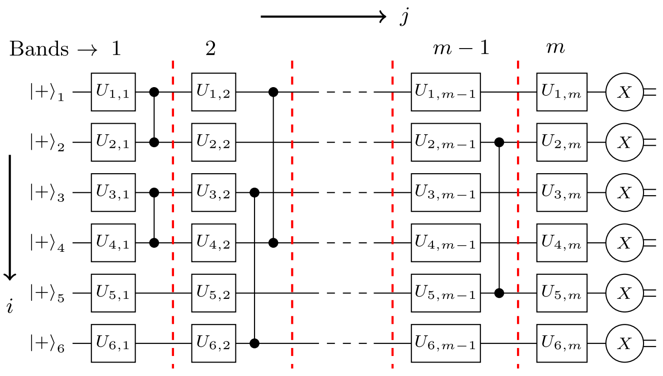

Figure 2. A six-qubit example of target circuit.

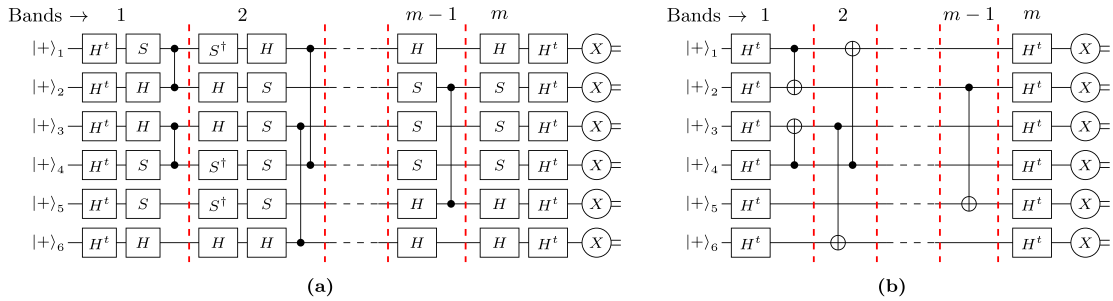

Figure 3. (a) Example of trap circuit for the target circuit in figure 2 and (b) overall computation implemented through this trap circuit. All the single-qubit gates acting on the same qubit in the same band must be recompiled into one gate—for instance, in figure 3(a), the H^t-gate and subsequent S-gate acting on qubit 1 in band 1 must be implemented as one gate SH^t.

sequence of bands, each one containing one round of single-qubit gates and one round of cZ gates. We will indicate the number of qubits with n and the number of bands with m.

In our accreditation protocol $v + 1$ circuits are implemented, one (chosen at random) being the target and the remaining v being the traps. The trap circuits are obtained by replacing the single-qubit gates in the target circuit with other single-qubit gates, but input state, measurements and cZ gates are the same as in the target (figure 3(a); all single-qubit gates acting on the same qubit in the same band must be recompiled into one gate). These single-qubit gates are chosen as follows (Routine 2, Box 3 in the Methods):

For each band $j \in \{1,\ldots,m-1\}$ and for each qubit $i \in \{1,\ldots,n\}$:

- If qubit i is connected to another qubit i' by a cZ gate, a gate is chosen at random from the set $\{H_i \otimes S_{i'},\ S_i \otimes H_{i'}\}$ and is implemented on qubits i and i' in band j. This gate is then undone in band $j+1$.
- Otherwise, if qubit i is not connected to any other qubit by a cZ gate, a gate is chosen at random from the set $\{H_i,\ S_i\}$ and is implemented on qubit i in band j. This gate is then undone in band $j+1$.

Moreover, depending on the random bit $t \in \{0, 1\}$, the traps may begin and end with a round of Hadamard gates. Since $(S \otimes H)cZ(S^\dagger \otimes H) = cX$, the trap circuits are a sequence of (randomly oriented) cX gates acting on $|+\rangle^{\otimes n}$ (if $t = 0$) or $|0\rangle^{\otimes n}$ (if $t = 1$)—figure 3(b). In the absence of noise, they always output $\bar{s} = \bar{0}$.

Our protocol requires appending a Quantum One-Time Pad (QOTP) to all single-qubit gates in all circuits (target and traps). This is described in Routine 1 in the Methods and is done as follows:

- For all bands $j = 1,\ldots,m$ and qubits $i = 1,\ldots,n$, a random Pauli gate is appended *after* each gate $U_{i,j}$ (figure 4(a)). This yields

$$U'_{i,j} = X_i^{\alpha'_{i,j}} Z_i^{\alpha_{i,j}} U_{i,j}, \tag{6}$$

where $\alpha_{i,j},\ \alpha'_{i,j} \in \{0, 1\}$ are random bits.

- For all bands $j = 2,\ldots,m$ and qubits $i = 1,\ldots,n$, another Pauli gate is appended *before* each single-qubit gate. This Pauli gate is chosen so that it undoes the QOTP coming from the previous band (figure 4(b)). Choosing

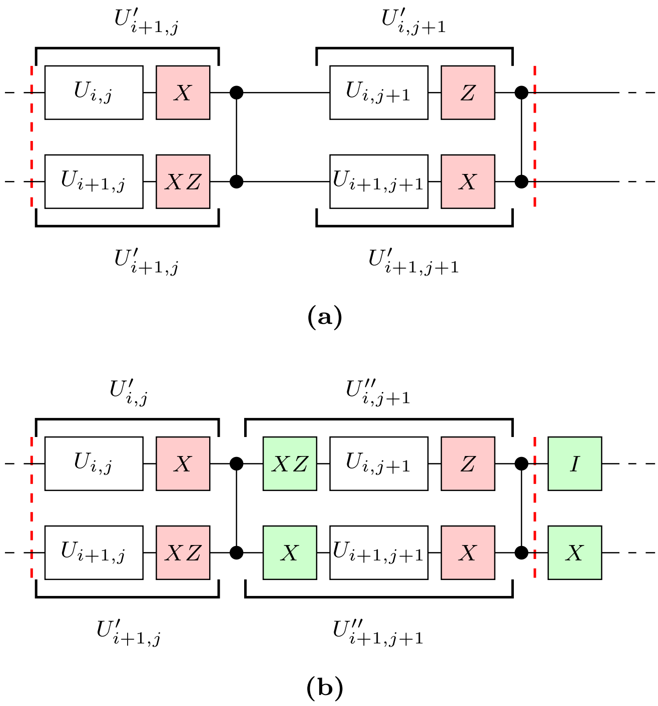

Figure 4. Example of Quantum one-time pad. (a) The red Pauli gates apply the QOTP. (b) The green gates undo the QOTP coming from previous bands.

this Pauli gate requires using the identities

$$(X_1 \otimes I_2)cZ = cZ(X_1 \otimes Z_2), \tag{7}$$

$$(Z_1 \otimes I_2)cZ = cZ(Z_1 \otimes I_2). \tag{8}$$

This yields

$$U''_{i,j} = X_i^{\alpha'_{i,j}} Z_i^{\alpha_{i,j}} U_{i,j} P_i(\{\alpha_{i,j-1}\}_i, \{\alpha'_{i,j-1}\}_i), \tag{9}$$

where $P_i(\{\alpha_{i,j-1}\}_i, \{\alpha'_{i,j-1}\}_i)$ is a Pauli gate that depends on the QOTP in the previous band.

- A random Pauli-X gate is appended *before* all the gates $U'_{i,1}$ in the first band. This yields

$$U''_{i,1} = U'_{i,1} X_i^{\gamma'_i} \tag{10}$$

with $\gamma_i \in \{0, 1\}$ chosen at random.

Overall, replacing each gate $U_{i,j}$ with $U''_{i,j}$ yields a new circuit that is equivalent to the the original one, apart from the un-recovered QOTP $\otimes_{i=1}^{n} X^{\alpha'_{i,m}} Z^{\alpha_{i,m}}$ in the last band. Since all measurements are in the Pauli-X basis, the Pauli-X component of this un-recovered QOTP is irrelevant, while its Pauli-Z component bit-flips some of the outputs. These bit-flips can be undone by replacing each output s_i with $s_i \oplus \alpha_{i,m}$ (a procedure that we call 'classical post-processing of the outputs'). This allows to recover the correct outputs.

After all the circuits have been implemented and the outputs have been post-processed, the flag bit is initialized to $|\text{acc}\rangle = |0\rangle$, then it is checked whether all the traps gave the correct output $\bar{s} = \bar{0}$. If they do, the protocol returns the output of the target together with the bit $|\text{acc}\rangle$, otherwise it returns the output of the target together with the bit $|\text{rej}\rangle = |1\rangle$. The output of the target is only accepted in the first case, while it is discarded in the second case.

In the absence of noise, our protocol always returns the correct output of the target circuit and always accepts it. Correctness of the target is ensured by the fact that the QOTP has no effect on the computation, as all the extra Pauli gates cancel out with each other or are countered by the classical post-processing of the outputs. Acceptance is ensured by the fact that in the absence of noise all the trap circuits always yield the correct outcome $\bar{s} = \bar{0}$.

We will now consider a noisy implementation of our protocol, explain the role played by the various tools (QOTP, trap circuits etc.) and show that with single-qubit gates suffering bounded noise, our protocol ensures that wrong outputs are rejected with high probability.

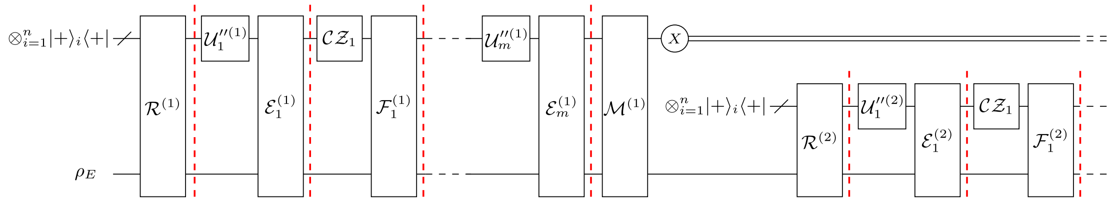

Figure 5. Schematic illustration of a noisy implementation of our protocol where all boxes represent CPTP maps. $\mathcal{U}_j^{\prime\prime(k)}$ implements the round of single-qubit gates in band j of circuit k, $\mathcal{CZ}_j$ implements the round of cZ gates in band j. In each circuit $k = 1,\ldots,v+1$: $\mathcal{R}^{(k)}$ is the noise in state preparation, $\mathcal{M}^{(k)}$ is the noise in measurements, $\mathcal{E}_j^{(k)}$ is the noise in the round of single-qubit gates in band j and $\mathcal{F}_j^{(k)}$ is the noise in the round of cZ gates in band j.

2.4. The credibility of our protocol

As per equation (2), we model noise as a set of CPTP maps acting on the whole system and on the environment (figure 5). For simplicity, let us begin with the assumption that all the rounds of single-qubit gates in our protocol are noiseless, i.e. that for all circuits $k = 1,\ldots,v+1$ and bands $j = 1,\ldots,m$, a noisy implementation of the round of single-qubit gates is (see figure 5 for notation)

$$\widetilde{\mathcal{U}}_j^{\prime\prime(k)} = \mathcal{E}_j^{(k)}(\mathcal{U}_j^{\prime\prime(k)} \otimes \mathcal{I}_E) \text{ with } \mathcal{E}_j^{(k)} = \mathcal{I}_{SE}, \tag{11}$$

where $\mathcal{I}_{SE}$ is the identity on system and environment. Under this assumption, a first simplification to the noise of type N1 comes from the QOTP, a tool used in many works in verification [25] and benchmarking protocols [20, 51] that also plays a crucial role in our protocol. If single-qubit gates are noiseless, the QOTP allows to randomize all noise processes, even those non-local in space and time, to classically correlated Pauli errors (see lemma 1 in appendix A). A similar result was previously proven in [51] for Markovian noise, and here we show that this result holds also if the noise creates correlations in time.

Having reduced arbitrary non-local noise to Pauli errors via the QOTP, we show (see lemma 2 in appendix B) that our trap circuits detect all Pauli errors with non-zero probability. The reasoning is as follows: since the trap circuits contain only Clifford gates, the noise acting at any point of a trap circuit can be factored to the end of the circuit. The noisy trap circuit is thus rewritten as the original one (figure 3(a)) with a Pauli-Z error $P \in \{I, Z\}^{\otimes n}$ applied before the measurements. If $P \neq I^{\otimes n}$, the trap outputs a wrong output ($\bar{s} \neq \bar{0}$) and the noise is detected. However, if the errors in different parts of the circuit happen to cancel out, then $P = I^{\otimes n}$, the trap outputs $\bar{s} = \bar{0}$ and the noise is not detected. The role of H and S-gates in our trap circuits is to ensure that this happens with suitably low probability for *all* types of noise that can possibly affect the trap. These gates map Pauli errors into other Pauli errors as

$$\begin{array}{ccccc} & & H & S & \\ X & \to & Z & Y & \\ Y & \to & Y & X & , \\ Z & \to & X & Z & \end{array} \tag{12}$$

where we omit unimportant prefactors (global phases do not affect outputs). Therefore, the random implementation of H and S-gates prevents errors in state preparation and two-qubit gates from canceling trivially. Similarly, the rounds of Hadamard gates activated at random at the beginning and at the end of the trap circuits prevent measurement errors from canceling trivially with noise happening before. These arguments are used to prove the claim of lemma 2, that states that our trap circuits can detect all possible Pauli errors with probability larger than 1/4.

The above arguments and lemmas can be used to prove that our protocol can detect arbitrary noise in state preparation, measurement and two-qubit gates, provided that single-qubit gates are noiseless:

Theorem 1. *Suppose that all single-qubit gates in our accreditation protocol are noiseless. For any number $v \geqslant 3$ of trap circuits, our accreditation protocol can accredit the outputs of a noisy quantum computer affected by noise of the form N1 with*

$$\varepsilon = \frac{\kappa}{v+1}, \tag{13}$$

where $\kappa = 3(3/4)^2 \approx 1.7$.

To prove $\varepsilon = \kappa/(v+1)$ we write the state of the system at the end of a noisy protocol run as in equation (3). We do this using lemmas 1 and 2. The proof of theorem 1 is in appendix C.

We now relax the assumption of noiseless single-qubit gates and generalize our results to noise of the form N2. We assume that all rounds of single-qubit gates suffer bounded noise, i.e. that for all circuits $k = 1,\ldots,v + 1$ and bands $j = 1,\ldots,m$, a noisy implementation of the round of single-qubit gates is (see Figure 5 for notation)

$$\widetilde{\mathcal{U}}''^{(k)}_j = \mathcal{E}^{(k)}_j(\mathcal{U}''^{(k)}_j \otimes \mathcal{I}_E) \tag{14}$$

with $\mathcal{E}^{(k)}_j = (1 - r^{(k)}_j)\mathcal{I}_{SE} + r^{(k)}_j\mathcal{E}'^{(k)}_j$ for some arbitrary CPTP map $\mathcal{E}'^{(k)}_j$ acting on both system and environment and for some number $0 \leqslant r^{(k)}_j < 1$. We refer to the number $r^{(k)}_j$ as 'error rate' of $\mathcal{U}''^{(k)}_j$. Since each $\mathcal{U}''^{(k)}_j$ is chosen at random (depending on whether circuit k is the target or a trap and on the QOTP) and since noise in single-qubit gates is potentially gate-dependent (condition N2), let us indicate with $r^{(k)}_{\max,j}$ the maximum error rate of single-qubit gates in band j of circuit k, the maximum being taken over all possible choices of $\mathcal{U}''^{(k)}_j$.

We can now state theorem 2:

Theorem 2. *Our protocol with $v \geqslant 3$ of trap circuits can accredit the outputs of a noisy quantum computer affected by noise of the form* N1 *and* N2 *with*

$$\varepsilon = g\frac{\kappa}{v + 1} + 1 - g, \tag{15}$$

where $\kappa = 3(3/4)^2 \approx 1.7$ and $g = \prod_{j,k}(1 - r^{(k)}_{\max,j})$.

To calculate ε for the protocol with noisy single-qubit gates we use that $\mathcal{E}^{(k)}_j = (1 - r^{(k)}_{\max,j})\mathcal{I}_{SE} + r^{(k)}_{\max,j}\mathcal{Q}^{(k)}_j$, where $\mathcal{Q}^{(k)}_j$ is a CPTP map encompassing the system and the environment. We can then rewrite the state of the system at the end of the protocol as

$$\rho^\star_{\text{out}} = g\rho_{\text{out}} + (1 - g)\,\widetilde{\rho}_{\text{out}}, \tag{16}$$

where ρ_{out} is the state of the system at the end of a protocol run with noiseless single-qubit gates—which by theorem 1 is of the form of equation (3) with $b \leqslant \kappa/(v + 1)$—and $\widetilde{\rho}_{\text{out}}$ is a quantum state containing the effects of noise in single-qubit gates. Expressing $\widetilde{\rho}_{\text{out}}$ as

$$\widetilde{\rho}_{\text{out}} = h\,\widetilde{\tau}^{\text{tar}}_1 \otimes |\text{acc}\rangle\langle\text{acc}| + (1 - h)\,\widetilde{\tau}^{\text{tar}}_2 \otimes |\text{rej}\rangle\langle\text{rej}|, \tag{17}$$

where $\widetilde{\tau}^{\text{tar}}_1$ and $\widetilde{\tau}^{\text{tar}}_2$ are arbitrary states for the target and $0 \leqslant h \leqslant 1$, we thus have

$$\begin{aligned}\rho^\star_{\text{out}} = g[b\,\tau'^{\text{tar}}_{\text{out}} \otimes |\text{acc}\rangle\langle\text{acc}| + (1 - b)(l\,\sigma^{\text{tar}}_{\text{out}} \otimes |\text{acc}\rangle\langle\text{acc}|\\ + (1 - l)\tau^{\text{tar}}_{\text{out}} \otimes |\text{rej}\rangle\langle\text{rej}|)] + (1 - g)[h\,\widetilde{\tau}^{\text{tar}}_1 \otimes |\text{acc}\rangle\langle\text{acc}|\\ + (1 - h)\,\widetilde{\tau}^{\text{tar}}_2 \otimes |\text{rej}\rangle\langle\text{rej}|].\end{aligned} \tag{18}$$

As it can be seen, the probability that the target is in the wrong state and the flag bit is in the state $|\text{acc}\rangle$ is $gb + (1 - g)h \leqslant g\kappa/(v + 1) + (1 - g)h$, where we used that $b \leqslant \kappa/(v + 1)$ from theorem 1. This probability reaches its maximum for $h = 1$, therefore we have $\varepsilon = g\kappa/(v + 1) + 1 - g$. Note that if $r^{(k)}_{\max,j} \leqslant r_0 \ll 1$, then

$$g \geqslant \prod_{k=1}^{v+1}\prod_{j=1}^{m}(1 - r_0) \approx 1 - m(v + 1)r_0 + O(r_0^2). \tag{19}$$

Thus, if $r_0 \ll 1/m(v + 1)$, then $g \approx 1$ and $\varepsilon \approx 1.7/(v + 1)$.

It is worth noting that our theorem 1 also holds if single-qubit gates suffer unbounded noise, provided that this noise is gate-independent. Indeed, if $\mathcal{E}^{(k)}_j = \mathcal{E}$ does not depend on the parameters in $\mathcal{U}''^{(k)}_j$ (see figure 5 for notation), using $\mathcal{E}^{(k)}_j\mathcal{CZ}_j\mathcal{F}^{(k)}_j = \mathcal{CZ}_j\mathcal{F}'^{(k)}_j$ (with $\mathcal{F}'^{(k)}_j = \mathcal{CZ}_j^{-1}\mathcal{E}^{(k)}_j\mathcal{CZ}_j\mathcal{F}^{(k)}_j$) we can factor this noise into that of $\mathcal{CZ}_j$ and prove $\varepsilon = \kappa/(v + 1)$ with the same arguments used in theorem 1. Similarly, we also expect our theorem 2 to hold if noise in single-qubit gates has a *weak* gate-dependence, as is the case for some of the protocols centered around randomized benchmarking [19]. We leave the analysis of weakly gate-dependent noise to future works.

2.5. Mesothetic verification protocol

In Box 4 in the Methods we translate our accreditation protocol into a cryptographic protocol, obtaining what we call 'mesothetic' verification protocol. To verify an n-qubit computation, in the mesothetic protocol the verifier (Alice) must possess a device that can receive n qubits from the prover (Bob), implement single-qubit gates on all of them and send the qubits back to the Bob. In appendix D we present theorems D1 and D2, which are the counterparts of theorems 1 and 2 for the cryptographic protocol. In the first two theorems the number ε

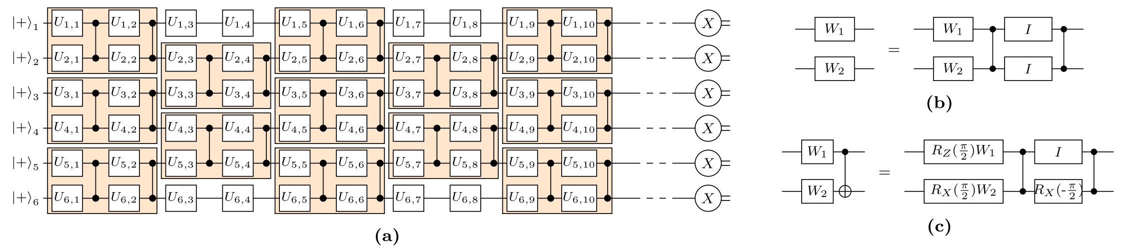

Figure 6. (a) Six-qubit example of circuit in normal form. This circuit has the same repetitive structure as the Brickwork States [52]. Recompiling the target circuit into a normal form of this type can always be done using the circuit identities (b) and (c).

is replaced by the soundness ε_{cr} (see definition 4 in appendix D.2), namely the probability that Alice accepts a wrong output for the target when Bob is cheating.

Our mesothetic verification protocol is different from prepare-and-send [26–33] or receive-and-measure [34–38] cryptographic protocols in that Alice encrypts the computation through the QOTP during the actual implementation of the circuits, and not before or after the implementation. To do this, she must possess an n-qubit quantum memory and be able to execute single-qubit gates. Despite being scalable, our protocol is more demanding that those in [27–38], where Alice only requires a single-qubit memory. This suggests the interesting possibility that protocols optimized for experiments may translate into more demanding cryptographic protocols and vice-versa.

Similarly to post-hoc verification protocols [35, 43], our protocol is not blind. Alice leaks crucial information to Bob regarding the target circuit, such as the position of two-qubit gates. This is not a concern for our goals, as verifiability in our protocol relies on Bob being incapable to distinguish between target and trap circuits, i.e. to retrieve the number v_0, see lemma D3 in appendix D.2.

Blindness may be required to protect user's privacy in future scenarios of delegated quantum computing [53]. In appendix D.2 we thus show how to make our protocol blind. This requires recompiling the target circuit into a circuit in normal form with fixed cZ gates, such as the brickwork-type circuit in figure 6. This yields an increase in circuit depth, hence the minimal overheads of our protocol must be traded for blindness.

3. Discussion

We have presented a trap-based accreditation protocol for the outputs of NISQ devices. Our protocol is scalable and practical, and relies on minimal assumptions on the noise. Specifically, our protocol requires that single-qubit gates suffer bounded (but potentially gate-dependent and non-local in space and time) noise.

A single protocol run ends by either accepting or rejecting the output of the target circuit that we seek to accredit. We can then run our accreditation protocol multiple times (with the same target and with the same number of traps), each time keeping the output if the protocol accepts and discarding it if the protocol rejects. After multiple runs with i.i.d. noise, our protocol allows to bound the variation distance between noisy and noiseless probability distribution of the accepted outputs (equation (1)).

Real-world devices can be accredited by running our protocol on them. The accreditation is provided by bounds on the variation distance that rely on ε, which we obtained in theorems 1 and 2, and the acceptance probability prob(acc) of our accreditation protocol. The latter is estimated experimentally by running our protocol multiple times on the device being accredited.

Some noise models allow to lower-bound the acceptance probability analytically and consequently to upper-bound the variation distance. For instance, if all operations $\mathcal{E}_S^{(p)}$ in our protocol suffer bounded noise and have error rates r_p, we can write the state of the system at the end of the protocol as

$$\rho_{\text{out}} = \delta\ \sigma_{\text{out}}^{\text{tar}} \otimes |\text{acc}\rangle\langle\text{acc}| + (1-\delta)\tilde{\tau}, \tag{21}$$

where $\sigma_{\text{out}}^{\text{tar}}$ is the state of the target at the end of a noiseless protocol run, $\tilde{\tau}$ is an arbitrary state for target and flag and $\delta = \prod_p (1-r_p) \in [0,\ 1]$. This yields prob(acc) $\geqslant \delta$ and (see equation (5))

$$\frac{1}{2}\sum_{\bar{s}} |p_{\text{noiseless}}(\bar{s}) - p_{\text{noisy}}(\bar{s})| \leqslant \frac{\varepsilon}{\text{prob(acc)}} \leqslant \frac{\varepsilon}{\delta}. \tag{22}$$

In figure 7 we plot the RHS of the above equation. Plots of this type can be used to seek error rates that will provide the desired upper-bound on the variation distance.

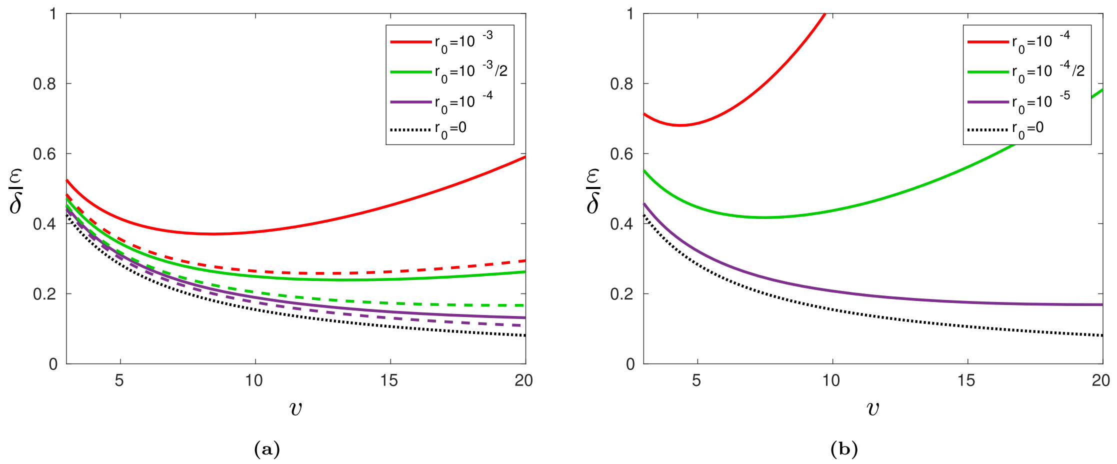

Figure 7. RHS of Inequality 22 for (a) target circuit preparing and measuring n-qubit GHZ states, with $n = m = 7$ (dashed lines) and $n = m = 10$ (solid lines) and (b) target circuit implementing a pseudo-random circuit for supremacy experiment [14] with $n = 62$ qubits and circuit depth $m = 34$. In these plots we assume that all operations are affected by bounded noise. We also assume that single-qubit state preparation, single-qubit measurements and cZ-gates have error rates r_0, and that all single-qubit gates have error rates $r_0/10$.

Considering the error rates of present NISQ devices [2, 45–48] we expect that our protocol may provide worthwhile upper-bounds for target circuits with up to $n \approx 7$ qubits and $m \approx 7$ bands (figure 7(a)). We also expect that larger target circuits may yield upper-bounds that are too large to be useful. In fact, for large target circuits, it is also possible that none of the protocol runs will accept the output of the target, and thus that our protocol will provide no upper-bound. Nevertheless, it is worth remarking that this does not indicate that our accreditation protocol is unable to accredit computations on NISQ devices. On the contrary, by providing large upper-bounds, our protocol reveals that the device being tested suffers high levels of noise and that its outputs are not credible.

Our work leaves several open questions. Our theorem 2 shows that our protocol requires reducing the error rates of single-qubit gates with the size of the target circuit. This requirement is similar to that found in other works [29, 38, 41], and is a known obstacle towards scalable quantum computing. A strategy that has been exploited in previous works is to incorporate fault-tolerance into the existing protocols [29, 41]. Another has been to define verifiable fault-tolerance using notions such as acceptability and detectability [54]. Interfacing fault tolerance with accreditation is an interesting challenge for the future.

Another open question regards the applicability of our accreditation protocol if single-qubit gates suffer unbounded noise. In its current state, the analysis of our protocol does not account for unbounded gate-dependent noise in single-qubit gates, including unitary errors such as over- or under-rotations. The reason is that the QOTP (which maps coherent errors into classically correlated Pauli errors) is applied at the level of single-qubit gates. Unbounded errors that depend on the gates used to randomize arbitrary noise processes to Pauli errors are an obstacle to other works including cryptographic protocols [29] and protocols based on randomized benchmarking [16–18, 20, 21].

Finally, with the mesothetic protocol we show how to adapt our protocol to the cryptographic setting. In the mesothetic protocol the verifier requires an n-qubit memory and the ability to execute single-qubit gates. This protocol ismore demanding than several existing cryptographic protocols [27–38] requiring single-qubit memory for the verifier. An interesting question is whether a mesothetic protocol can be devised that only requires single-qubit gates and single-qubit memory for the verifier.

4. Methods

4.1. Overhead of our accreditation protocol

Here we count the overhead of our protocol. Our protocol has no quantum overhead, as all circuits have the same size as the one being verified. The classical overhead consists in O(nm) bits for each of the $v + 1$ computations. Specifically, the target computation has an overhead of $2nm + n$ bits (the $2nm$ random bits $\alpha_{i,j}$, $\alpha'_{i,j}$ and the n random bits γ_i in Routine 1), while the traps have an overhead of at most $2nm + n + nm$ bits (the $2nm + n$ random bits in Routine 1 and at most nm random bits in Routine 2).

Box 1. Accreditation protocol.

Input:

1. A target circuit that takes as input n qubits in the state $|+\rangle$, contains only single-qubit gates and cZ gates arranged in m bands and ends with Pauli-X measurements (figure 2).
2. The number v of trap circuits.

Routine:

1. Choose a random number $v_0 \in \{1,\ldots,v+1\}$ and define $\{U_{i,j}^{(v_0)}\} = \{U_{i,j}\}$, where $\{U_{i,j}\}$ is the set of single-qubit gates in the target circuit.
2. For $k = 1,\ldots,v+1$: If $k \neq v_0$ (trap circuit), run Routine 2 and obtain the set of single-qubit gates $\{U_{i,j}^{(k)}\}$ for the kth trap circuit.
3. For $k = 1,\ldots,v+1$: run Routine 1 and obtain $\{U_{i,j}^{\prime\prime(k)}\}$, together with the bit-string $(\alpha_{1,m}^{(k)},\ldots\alpha_{n,m}^{(k)})$.
4. For $k = 1,\ldots,v+1$:
   4.1 Create a state $\rho_{\text{in}} = |+\rangle^{\otimes n}$.
   4.2 Implement circuit k with single-qubit gates from the set $\{U_{i,j}^{\prime\prime(k)}\}$ and obtain output $\bar{s}^{(k)} = (s_1^{(k)},\ldots,s_n^{(k)})$. Next, for all $i = 1,\ldots,n$, recompute $s_i^{(k)}$ as $s_i^{(k)} \oplus \alpha_{i,m}^{(k)}$.
5. Initialize a flag bit to $|\text{acc}\rangle = |0\rangle$. Then, for $k = 1,\ldots,v+1$: if $\bar{s}^{(k)} \neq \bar{0}$ and $k \neq v_0$ (trap circuit), set the flag bit to $|\text{rej}\rangle = |\text{acc} \oplus 1\rangle$.

Output:

The output $\bar{s}^{(v_0)}$ of the target circuit and the flag bit.

Box 2. Routine 1. (Quantum one-time pad).

Input:

A set $\{U_{i,j}\}$ of single-qubit gates, for $j = 1,\ldots,m$ and $i = 1,\ldots,n$.

Routine:

1. For $j = 1,\ldots,m$ and $i = 1,\ldots,n$:
Choose two random bits $\alpha_{i,j}$ and $\alpha'_{i,j}$. Next, define $U'_{i,j} = X^{\alpha'_{i,j}} Z^{\alpha_{i,j}} U_{i,j}$.

2. For $i = 1,\ldots,n$:
Choose a random bit γ_i and define $U''_{i,1} = U'_{i,1} X^{\gamma_i}$.

3. For $j = 1,\ldots,m-1$:
Using equations (7) and (8) define $U''_{i,j+1}$ so that

$$(\otimes_i\, U''_{i,j+1})\widehat{cZ}_j(\otimes_i\, U'_{i,j}) = (\otimes_i\, U'_{i,j+1})\widehat{cZ}_j(\otimes_i\, U_{i,j}), \tag{20}$$

where $\widehat{cZ}_j$ is the entangling operation in the jth band.

Output:

The set $\{U''_{i,j}\}$ and the n-bit string $(\alpha_{1,m},\ldots,\alpha_{n,m})$.

Box 3. Routine 2. (Single-qubit gates for trap circuits).

Input:

The target circuit.

Routine:

1. Initialize the set $\{U_{i,j} = I_i\}$, for $i = 1,\ldots,n$ and $j = 1,\ldots,m$.
2. For all $j = 1,\ldots,m-1$:
   2.1 For all $i = 1,\ldots,n$: If in band j of the target circuit qubits i and i' are connected by a cZ gate, set
   • $U_{i,j} = S_i U_{i,j}$ and $U_{i',j} = H_{i'} U_{i',j}$ with probability 1/2.
   • $U_{i,j} = H_i U_{i,j}$ and $U_{i',j} = S_{i'} U_{i',j}$ with probability 1/2.
   Otherwise, set $U_{i,j} = S_i U_{i,j}$ or $U_{i,j} = H_i U_{i,j}$ with probability 1/2.
   2.2 For all $i = 1,\ldots,n$: Set $U_{i,j+1} = U_{i,j}^{\dagger}$.
3. For all $i = 1,\ldots,n$:
Choose a random bit $t \in \{0, 1\}$. Next, set $U_{i,1} = U_{i,1} H^t$ and $U_{i,m} = H^t U_{i,m}$.

Output:

The set $\{U_{i,j}\}$.

Box 4. Mesothetic protocol (further details in appendix D.)

Input:
A classical description of the target circuit and the number v of traps. (The input is known to both Alice and Bob).

Preliminary Operations:
1. Alice randomly chooses which circuit $v_0 \in \{1,\ldots,v+1\}$ will be used to implement the target. Next she defines $\{U_{i,j}^{(v_0)}\} = \{U_{i,j}\}$, where $\{U_{i,j}\}$ is the set of single-qubit gates in the target circuit.
2. For $k = 1,\ldots,v+1$: If $k \neq v_0$ (trap circuit), Alice runs Routine 2 and obtains the set $\{U_{i,j}^{(k)}\}$ of single-qubit gates for the kth circuit.
3. For $k = 1,\ldots,v+1$: Alice runs Routine 1 and obtains the set of gates $\{U_{i,j}^{\prime\prime(k)}\}$, together with the random bits $\alpha_{i,m}^{(k)}$.

Routine:
4. For all $k = 1,\ldots,v+1$, Alice and Bob interact as follows:
  4.1 Bob creates n qubits in state $|+\rangle$.
  4.2 For $j = 1,\ldots,m$:
    4.2.1 Bob sends all the qubits to Alice. For $i = 1,\ldots,n$, Alice executes $U_{i,j}^{\prime\prime(k)}$ on qubit i. Finally, Alice sends all the qubits back to Bob.
    4.2.2 Bob applies the entangling gates $\widehat{cZ}_j$ contained in the jth band of the target circuit.
  4.3 For $i = 1,\ldots,n$: Bob measures qubit i in the Pauli-X basis and communicates the output to Alice. Alice bit-flips the output if $\alpha_{i,m}^{(k)} = 1$, otherwise she does nothing. Next, if $k \neq v_0$ and this output is $s_i = 1$, Alice aborts the protocol, otherwise she does nothing.
5. Alice initializes a flag bit to the state $|\mathrm{acc}\rangle = |0\rangle$. Next, for all $k = 1,\ldots,v+1$: if $k \neq v_0$ (trap circuit) and $s_i^{(k)} \oplus \alpha_{i,m}^{(k)} \neq 0$ for some $i \in \{1,\ldots,n\}$, Alice sets the flag bit to $|\mathrm{rej}\rangle = |\mathrm{acc} \oplus 1\rangle$.

Output:
The outputs of the target circuit and the flag bit.

Acknowledgments

This research was supported by the UK EPSRC (EP/K04057X/2) and the UK Networked Quantum Information Technologies (NQIT) Hub (EP/M013243/1). We acknowledge helpful discussions with Joel Wallman, Elham Kashefi, Dominic Branford, Zhang Jiang, Joseph Emerson.

Notation

In these appendices we will indicate the action of the round of single-qubit gates in a band $j \in \{1,\ldots,m\}$ of a circuit $k \in \{1,\ldots,v+1\}$ as $\mathcal{U}_j^{\prime\prime(k)}(\rho_S) = \otimes_{i=1}^n U_{i,j}^{\prime\prime(k)}(\rho_S) U_{i,j}^{\dagger\prime\prime(k)}$, where ρ_S is the state of the system. Similarly, we will indicate the action of a round of cZ gates on the system as $\mathcal{CZ}_j(\rho_S) = \widehat{cZ}_j(\rho_S)\widehat{cZ}_j$, where $\widehat{cZ}_j$ is the tensor product of all cZ gates in band j in the target circuit, and the action of n-qubit Pauli operators as $\mathcal{P} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes n}$, where $\mathcal{I}(\rho) = \rho$, $\mathcal{X}(\rho) = X\rho X$, $\mathcal{Y}(\rho) = Y\rho Y$, $\mathcal{Z}(\rho) = Z\rho Z$ are single-qubit Pauli operators.

In appendix A we provide statement and proof of lemma 1, In appendix B we provide statement and proof of lemma 2 , In appendix C we prove theorem 1 and in appendix D we prove soundness of the mesothetic protocol.

Appendix A. Statement and proof of lemma 1

We now present and prove lemma 1, which is as follows:

Lemma 1. *Suppose that all single-qubit gates in all the circuits implemented in our protocol are noiseless, and that state preparation, measurements and two-qubit gates suffer noise of the type type N1. For a fixed choice of single-qubit gates $\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}$, summed over all the random bits $\alpha_{i,j}$, $\alpha'_{i,j}$, γ_i (see Routine 1), the joint state of the target circuit and of the traps after they have all been implemented is of the form*

$$\rho_{\mathrm{out}}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}) = \sum_{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}} \sum_{\mathcal{P}_0^{(1)},\ldots,\mathcal{P}_m^{(v+1)}} \frac{\mathrm{prob}(\mathcal{P}_0^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}} \times \bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)} \circ_{j=1}^{m-1}(\mathcal{CZ}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)}) \circ \mathcal{P}_0^{(k)}(\rho_{\mathrm{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}), \tag{A1}$$

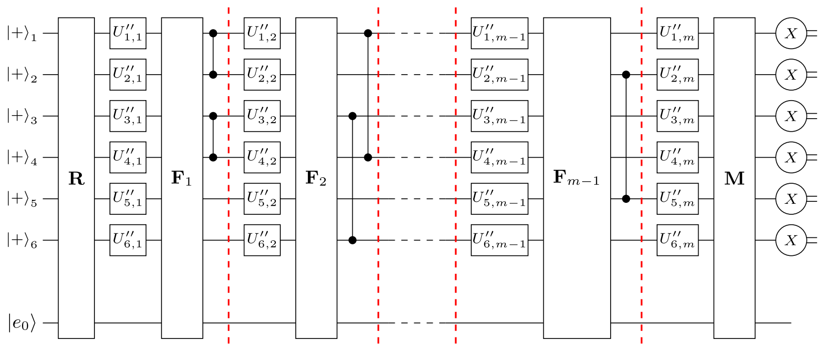

Figure 8. Noisy implementation of the 6-qubit target circuit in figure 2. The noise in state preparation is described by the unitary $\mathbf{R}$, that in the measurements by $\mathbf{M}$, that in the cZ-gates in a band $j = 1,\ldots,m-1$ by $\mathbf{F}_j$. All these unitaries act simultaneously on the system and on the environment (initially in the ground state $|e_0\rangle$).

where $\rho_{\rm in} = \otimes_i |+\rangle_i\langle +|$, $\bar{s}^{(k)} = (s_1^{(k)},\ldots,s_n^{(k)})$ *is a binary string representing the output of the kth circuit,* $\mathcal{Z}^{\bar{s}^{(k)}}(\rho) = \otimes_i Z_i^{s_i^{(k)}}\rho Z_i^{s_i^{(k)}}$ *and* $\mathrm{prob}(\mathcal{P}_0^{(1)},\ldots,\mathcal{P}_m^{(v+1)})$ *is the joint probability of a collection of Pauli errors* $\mathcal{P}_0^{(1)},\ldots,\mathcal{P}_m^{(v+1)}$ *affecting the system, with* $\mathcal{P}_1^{(k)},\ldots,\mathcal{P}_{m-1}^{(k)} \in \{\mathcal{I},\mathcal{X},\mathcal{Y},\mathcal{Z}\}^{\otimes n}$ *and* $\mathcal{P}_0^{(k)}$, $\mathcal{P}_m^{(k)} \in \{I,\mathcal{Z}\}^{\otimes n}$ *for all k.*

This lemma shows that if single-qubit gates are noiseless, the QOTP allows to reduce noise of the type N1 to classically correlated Pauli errors. These Pauli errors affect each circuit after state preparation ($\mathcal{P}_0^{(k)}$), before each entangling operation $\mathcal{CZ}_j$ ($\mathcal{P}_j^{(k)}$, for $j \in \{1,\ldots,m-1\}$) and before the measurements ($\mathcal{P}_m^{(k)}$). Errors in the cZ gates can be Pauli-X, Y and Z, while those in state preparation and measurements are Pauli-Z (this is because their Pauli-X components stabilize $\rho_{\rm in}$ and Pauli-X measurements respectively).

The main tool used in this section is the 'Pauli Twirl' [18].

(Pauli Twirl). Let ρ be a $2^n \times 2^n$ density matrix and let P, P' be two n-fold tensor products of the set of Pauli operators $\{I, X, Y, Z\}$. Denoting by $\{Q_r\}$ the set of all n-fold tensor products of the set of Pauli operators $\{I, X, Y, Z\}$,

$$\sum_{r=1}^{4^n} Q_r P Q_r \rho Q_r P' Q_r = 0\ \forall P \neq P'. \tag{A2}$$

We will also use a restricted version of the Pauli Twirl, which is proven in [29]

(Restricted Pauli Twirl). Let ρ be a $2^n \times 2^n$ density matrix and let P, P' be two n-fold tensor products of the set of Pauli operators $\{I, Z\}$. Denoting by $\{Q_r\}$ the set of all n-fold tensor products of the set of Pauli operators $\{I, X\}$,

$$\sum_{r=1}^{2^n} Q_r P Q_r \rho Q_r P' Q_r = 0\ \forall P \neq P'. \tag{A3}$$

The same holds if P and P' are two n-fold tensor products of the set of Pauli operators $\{I, X\}$ and $\{Q_r\}$ is the set of all n-fold tensor products of the set of Pauli operators $\{I, Z\}$.

Proof. *(Lemma 1)* We start proving the lemma for the case where we run a single circuit ($v = 0$), and then we generalize to multiple circuits ($v > 0$). Including all purifications in the environment, we can rewrite the noise as unitary matrices acting on system and environment (for clarity we write these unitaries in bold font). Doing this, for a fixed choice of gates $\mathcal{U}''_j,\ldots,\mathcal{U}''_m$ (which depend on the choice of gates $\mathcal{U}_1,\ldots,\mathcal{U}_m$ and on all the random bits $\alpha_{i,j}$, $\alpha'_{i,j}$, γ_i, see Routine 1), the state of the system before the measurement becomes (figure 8)

$$\rho(\mathcal{U}''_1,\ldots,\mathcal{U}''_m) = \mathrm{Tr}_E[\mathbf{M}\ U''_m\widehat{cZ}_{m-1}\mathbf{F}_{m-1}U''_{m-1}\ \ldots\ \widehat{cZ}_1\mathbf{F}_1U''_1\mathbf{R}(\rho_{\rm in}\otimes\rho_E)\mathbf{R}^\dagger\ U_1''^\dagger\mathbf{F}_1^\dagger\widehat{cZ}_1\ \ldots\ U''^\dagger_{m-1}\mathbf{F}^\dagger_{m-1}\widehat{cZ}_{m-1}U''^\dagger_m\mathbf{M}^\dagger], \tag{A4}$$

where $\rho_{\rm in} = \otimes_i |+\rangle_i\langle +|$, $\rho_E = |e_0\rangle\langle e_0|$ is the initial state of the environment, $U''_j = \otimes_i U''_{i,j}$ are the gates output by Routine 1, the unitary matrix $\mathbf{R}$ represents the noise in state preparation, the unitary matrix $\mathbf{M}$ represents the noise in the measurements, $\widehat{cZ}_j\mathbf{F}_j$ is the noisy round of entangling gates in a band j and $\mathrm{Tr}_E[\cdot]$ is the trace over the environment.

Table 1. Operators appearing in equation (A6).

$\langle e_{k_2}\lvert\mathbf{R}\rvert e_0\rangle = \sum_{\mu_1}\eta^{(R)}_{k_2,0,\mu_1}P_{\mu_1}$	$\langle e_{k_1}\lvert\mathbf{F}_1\rvert e_{k_2}\rangle = \sum_{\mu_2}\eta^{(F)}_{k_1,k_2,\mu_2}P_{\mu_2}$	$\langle e_p\lvert\mathbf{M}\rvert e_{k_1}\rangle = \sum_{\mu_3}\eta^{(M)}_{p,k_1,\mu_3}P_{\mu_3}$
$\langle e_0\lvert\mathbf{R}^\dagger\rvert e_{l_2}\rangle = \sum_{\nu_1}\eta^{*(R)}_{l_2,0,\nu_1}P_{\nu_1}$	$\langle e_{l_2}\lvert\mathbf{F}_1^\dagger\rvert e_{l_1}\rangle = \sum_{\nu_2}\eta^{*(F)}_{l_1,l_2,\nu_2}P_{\nu_2}$	$\langle e_{l_1}\lvert\mathbf{M}^\dagger\rvert e_p\rangle = \sum_{\nu_3}\eta^{*(M)}_{p,l_1,\nu_3}P_{\nu_3}$

For simplicity, we first prove our result for a circuit with $m = 2$ bands and generalize to $m > 2$ bands later. In this case, defining an orthonormal basis $\{|e_p\rangle\langle e_p|\}$ for the environment, the state in equation (A4) is

$$\rho(\mathcal{U}_1'', \mathcal{U}_2'') = \sum_p \langle e_p|\mathbf{M}\ U_2''\widehat{cZ}_1\mathbf{F}_1 U_1''\mathbf{R}(\rho_{\text{in}}\otimes|e_0\rangle\langle e_0|)\mathbf{R}^\dagger\ U_1''^\dagger\mathbf{F}_1^\dagger\widehat{cZ}_1 U_2''^\dagger\mathbf{M}^\dagger|e_p\rangle. \tag{A5}$$

Introducing resolutions of the identity on the environment before and after every noise operator, we have

$$\rho(\mathcal{U}_1'', \mathcal{U}_2'') = \sum_{\substack{p\\ k_1,k_2,l_1,l_2}} [\langle e_p|\mathbf{M}|e_{k_1}\rangle\ U_2''\widehat{cZ}_1\langle e_{k_1}|\mathbf{F}_1|e_{k_2}\rangle\, U_1''\langle e_{k_2}|\mathbf{R}|e_0\rangle](\rho_{\text{in}}) \times [\langle e_0|\mathbf{R}^\dagger|e_{l_2}\rangle\ U_1''^\dagger\langle e_{l_2}|\mathbf{F}_1^\dagger|e_{l_1}\rangle\widehat{cZ}_j U_2''^\dagger\langle e_{l_1}|\mathbf{M}^\dagger|e_p\rangle], \tag{A6}$$

since $\sum_k|e_k\rangle\langle e_k| = I_E$ and $\langle e_k|V_S|e_{k'}\rangle = V_S\delta_{k,k'}$ for every operator V_S acting only on the system. The operators $\langle e_p|\mathbf{M}|e_{k_1}\rangle$, $\langle e_{k_1}|\mathbf{F}_1|e_{k_2}\rangle$, $\langle e_{k_2}|\mathbf{R}|e_0\rangle$, $\langle e_0|\mathbf{R}^\dagger|e_{l_2}\rangle$, $\langle e_{l_2}|\mathbf{F}_1^\dagger|e_{l_1}\rangle$, $\langle e_{l_1}|\mathbf{M}^\dagger|e_p\rangle$ act only on the system, and can thus be written as in table 1.

In table 1, $P_\mu\in\{I, X, Y, Z\}^{\otimes n}$ are n-fold tensor products of Pauli operators acting on the system and $\eta^{(R)}_{k_2,0,\mu_1}, \eta^{(F)}_{k_1,k_2,\mu_2}, \eta^{(M)}_{k_p,k_1,\mu_3}$ are complex numbers. We then obtain

$$\rho(\mathcal{U}_1'', \mathcal{U}_2'') = \sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ \nu_1,\nu_2,\nu_3}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\nu_1}\eta^{*(F)}_{l_1,l_2,\nu_2}\eta^{*(M)}_{p,l_1,\nu_3})\ P_{\mu_3}U_2''\widehat{cZ}_1 P_{\mu_2}U_1''P_{\mu_1}(\rho_{\text{in}})P_{\nu_1}U_1''^\dagger P_{\nu_2}\widehat{cZ}_1 U_2''^\dagger P_{\nu_3}. \tag{A7}$$

We will now describe how to apply the Pauli twirl lemmas iteratively, in the order the operations apply on the input. Therefore, we start by showing how to eliminate terms of the sum where $\mu_1\neq\nu_1$. Since X stabilizes $|+\rangle$ states, we can rewrite ρ_{in} as $(\otimes_i X_i^{\gamma_i})\rho_{\text{in}}(\otimes_i X_i^{\gamma_i})$. Moreover, using $U_1'' = U_1'(\otimes_i X_i^{\gamma_i})$, see Routine 1, the above state becomes

$$\rho(\mathcal{U}_1'', \mathcal{U}_2'') = \sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ \nu_1,\nu_2,\nu_3}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\nu_1}\eta^{*(F)}_{l_1,l_2,\nu_2}\eta^{*(M)}_{p,l_1,\nu_3}) P_{\mu_3}U_2''\widehat{cZ}_1 P_{\mu_2}U_1'(\otimes_i X_i^{\gamma_i})P_{\mu_1}(\otimes_i X_i^{\gamma_i})(\rho_{\text{in}})(\otimes_i X_i^{\gamma_i})P_{\nu_1}(\otimes_i X_i^{\gamma_i})U_1'^\dagger P_{\nu_2}\widehat{cZ}_1 U_2''^\dagger P_{\nu_3}. \tag{A8}$$

Summing over all possible γ_i and applying the Restricted Pauli Twirl (the Pauli-X components of both P_{μ_1} and P_{ν_1} stabilize ρ_{in} and can thus be ignored), we obtain a factor δ_{μ_1,ν_1}, and thus the above state becomes

$$\begin{aligned}\rho(\mathcal{U}_1', \mathcal{U}_2'') &= \frac{1}{2^n}\sum_{\{\gamma_i\}}\rho(\mathcal{U}_1'', \mathcal{U}_2'')\\ &= \sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ \nu_2,\nu_3}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\nu_2}\eta^{*(M)}_{p,l_1,\nu_3})P_{\mu_3}U_2''\widehat{cZ}_1 P_{\mu_2}U_1'P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1'^\dagger P_{\nu_2}\widehat{cZ}_1 U_2''^\dagger P_{\nu_3}.\end{aligned} \tag{A9}$$

To operate a Pauli twirl on P_{μ_2} and P_{ν_2}, we rewrite U_1' as $(\otimes_i Z^{\alpha_{i,1}}X^{\alpha'_{i,1}})U_1$ and $U_2''\widehat{cZ}_1$ as $U_2'\widehat{cZ}_1(\otimes_i X^{\alpha'_{i,1}}Z^{\alpha_{i,1}})$, see Routine 1. Summing over $\alpha_{i,1}$ and $\alpha'_{i,1}$ and using the Pauli Twirl, we obtain δ_{μ_2,ν_2}, and thus

$$\begin{aligned}\rho(\mathcal{U}_1, \mathcal{U}_2') &= \frac{1}{2^{2n}}\sum_{\{\alpha_{i,1}\},\{\alpha'_{i,1}\}}\rho(\mathcal{U}_1', \mathcal{U}_2'')\\ &= \sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ \nu_3}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\nu_3})P_{\mu_3}U_2'\widehat{cZ}_1 P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1 U_2'^\dagger P_{\nu_3}.\end{aligned} \tag{A10}$$

To operate a Pauli twirl on P_{μ_3} and P_{ν_3} we write the state of the system after the measurements:

$$\begin{aligned}
\rho_{\text{meas}}(\mathcal{U}_1, \mathcal{U}_2') &= \frac{1}{2^n}\sum_{\{s_i\}} \otimes_i \ (\langle +|_i Z_i^{s_i}\ \rho(\mathcal{U}_1, \mathcal{U}_2')\ Z_i^{s_i}|+\rangle_i)\ Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i} \\
&= \frac{1}{2^n}\sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3,\nu_3\\ s_1,\dots,s_n}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\nu_3}) \\
&\times\langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}U_2'\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2'^\dagger P_{\nu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n}\ \times\ (\otimes_i Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i}) \\
&= \frac{1}{2^n}\sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3,\nu_3\\ s_1,\dots,s_n}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\nu_3}) \\
&\times\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}(\otimes_i X_i^{\alpha_{i,2}'}Z_i^{\alpha_{i,2}})U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger(\otimes_i X_i^{\alpha_{i,2}'}Z_i^{\alpha_{i,2}})P_{\nu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n} \\
&\times(\otimes_i Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i}),
\end{aligned}\tag{A11}$$

where in the second equality we used $U_2' = (\otimes_i X_i^{\alpha_{i,2}'}Z_i^{\alpha_{i,2}})U_2$. We can now rewrite $|+\rangle^{\otimes n}$ as $\otimes_i X_i^{\alpha_{i,2}''}|+\rangle^{\otimes n}$ and sum over $\{\alpha_{i,2}'\}$. Using the Restricted Pauli Twirl (the Pauli-X components of both P_{μ_3} and P_{ν_3} stabilize $|+\rangle^{\otimes n}$ and can thus be ignored), we obtain δ_{μ_3,ν_3}:

$$\begin{aligned}
\rho_{\text{meas}}(\mathcal{U}_1, \mathcal{U}_2) &= \frac{1}{2^n}\sum_{\alpha_{i,2}'} \rho_{\text{meas}}(\mathcal{U}_1, \mathcal{U}_2') \\
&= \frac{1}{2^n}\sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3,\nu_3\\ s_1,\dots,s_n}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\mu_3}) \\
&\times\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}(\otimes_i Z_i^{\alpha_{i,2}})U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger(\otimes_i Z_i^{\alpha_{i,2}})P_{\mu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n} \\
&\times\ \otimes_i\ \ Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i}.
\end{aligned}\tag{A12}$$

Finally, after the classical post-processing (which replaces the outputs s_i with $s_i \oplus \alpha_{i,2}$), average over $\{\alpha_{i,2}\}$ yields the outcome state

$$\begin{aligned}
\rho_{\text{out}}(\mathcal{U}_1,\ \mathcal{U}_2) &= \frac{1}{2^n}\sum_{\{\alpha_{i,2}\}}\frac{1}{2^n}\sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ s_1,\dots,s_n}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\mu_3}) \\
&\times\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}(\otimes_i Z_i^{\alpha_{i,2}})U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger(\otimes_i Z_i^{\alpha_{i,2}})P_{\mu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n} \\
&\times\ \otimes_i\ \ Z_i^{s_i\oplus\alpha_{i,2}}|+\rangle_i\langle +|Z_i^{s_i\oplus\alpha_{i,2}} \\
&= \frac{1}{2^n}\sum_{\{\alpha_{i,2}\}}\frac{1}{2^n}\sum_{\substack{p,k_1,k_2,l_1,l_2\\ \mu_1,\mu_2,\mu_3\\ s_1,\dots,s_n}} (\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\mu_3}) \\
&\times\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i\oplus\alpha_{i,2}})P_{\mu_3}(\otimes_i Z_i^{\alpha_{i,2}})U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger(\otimes_i Z_i^{\alpha_{i,2}})P_{\mu_3} \\
&\times\ (\otimes_i Z_i^{s_i\oplus\alpha_{i,2}})|+\rangle^{\otimes n} \\
&\times\ \otimes_i\ \ Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i} \\
&= \frac{1}{2^n}\sum_{\substack{\mu_1,\mu_2,\mu_3\\ s_1,\dots,s_n}}\left(\sum_{p,k_1,k_2,l_1,l_2}\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\mu_3}\right) \\
&\times\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger P_{\mu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n} \\
&\times\ \otimes_i\ \ Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i} \\
&= \frac{1}{2^n}\sum_{\substack{\mu_1,\mu_2,\mu_3\\ s_1,\dots,s_n}}\phi_{\mu_1,\mu_2,\mu_3}\ \langle +|^{\otimes n}(\otimes_i Z_i^{s_i})P_{\mu_3}U_2\widehat{cZ}_1P_{\mu_2}U_1P_{\mu_1}(\rho_{\text{in}})P_{\mu_1}U_1^\dagger P_{\mu_2}\widehat{cZ}_1U_2^\dagger P_{\mu_3}(\otimes_i Z_i^{s_i})|+\rangle^{\otimes n} \\
&\times\ \otimes_i\ \ Z_i^{s_i}|+\rangle_i\langle +|Z_i^{s_i},
\end{aligned}\tag{A13}$$

where

$$\phi_{\mu_1,\mu_2,\mu_3} = \sum_{p,k_1,k_2,l_1,l_2} \eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\eta^{*(R)}_{l_2,0,\mu_1}\eta^{*(F)}_{l_1,l_2,\mu_2}\eta^{*(M)}_{p,l_1,\mu_3} = \sum_p \left|\sum_{k_1,k_2}\eta^{(R)}_{k_2,0,\mu_1}\eta^{(F)}_{k_1,k_2,\mu_2}\eta^{(M)}_{p,k_1,\mu_3}\right|^2 \geqslant 0 \quad \text{(A14)}$$

and $\sum_{\mu_1,\mu_2,\mu_3}\phi_{\mu_1,\mu_2,\mu_3} = 1$. $\rho_{\text{out}}(\mathcal{U}_1, \mathcal{U}_2)$ is therefore a convex combination of quantum states and ϕ_{μ_1,μ_2,μ_3} can be seen as the joint probability of Pauli errors P_{μ_1}, P_{μ_2} and P_{μ_3}. We can thus rewrite

$$\rho_{\text{out}}(\mathcal{U}_1, \mathcal{U}_2) = \sum_{\substack{\mathcal{P}_0,\mathcal{P}_1,\mathcal{P}_2\\ s_1,\ldots,s_n}} \frac{\text{prob}(\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2)}{2^n} \langle+|^{\otimes n}[\mathcal{Z}^{\bar{s}}\mathcal{P}_2\,\mathcal{U}_2\,\mathcal{CZ}_1\,\mathcal{P}_1\,\mathcal{U}_1\,\mathcal{P}_0(\rho_{\text{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i}|+\rangle_i\langle+|Z_i^{s_i}), \quad \text{(A15)}$$

where $\mathcal{P}_0, \mathcal{P}_2 \in \{\mathcal{I}, \mathcal{Z}\}^{\otimes n}$, $\mathcal{P}_1 \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes n}$ and prob$(\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2)$ is the joint probability of Pauli errors $\mathcal{P}_0, \mathcal{P}_1, \mathcal{P}_2$. This concludes the proof for the protocol with $v = 0$ and $m = 2$.

The generalization to a protocol with $v = 0$ and $m > 2$ is straightforward. Starting from the state in equation (A4), one can use the same arguments as for the two-band circuit. To generalize to multiple circuits ($v > 0$), we start by noticing that the circuits are implemented in series, hence the noise can only affect one circuit at a time. By the principle of deferred measurements, we can execute all the measurements at the end of the protocol. Moreover, we can prepare the input qubits for all the circuits at the beginning of the protocol. Doing this, the state of the system after all circuits have been implemented becomes

$$\begin{aligned}\rho(\mathcal{U}_1''^{(1)},\ldots,\mathcal{U}_m''^{(v+1)}) = \text{Tr}_E[&(\mathbf{M}^{(v+1)}\,U_m''^{(v+1)}\widehat{cZ}_{m-1}\mathbf{F}^{(v+1)}_{m-1}U_{m-1}''^{(v+1)}\ldots\widehat{cZ}_1\mathbf{F}^{(1)}_{v+1}U_1''^{(v+1)}\mathbf{R}^{(v+1)})\ldots\\ &(\mathbf{M}^{(1)}\,U_m''^{(1)}\widehat{cZ}_{m-1}\mathbf{F}^{(1)}_{m-1}U_{m-1}''^{(1)}\ldots\widehat{cZ}_1\mathbf{F}^{(1)}_1U_1''^{(1)}\mathbf{R}^{(1)})\\ &(\rho_{\text{in}}\otimes\rho_E)(\mathbf{R}^{\dagger(1)}U_1''^{\dagger(1)}\mathbf{F}_1^{\dagger(1)}\widehat{cZ}_1\ldots U_{m-1}''^{\dagger(1)}\mathbf{F}^{\dagger(1)}_{m-1}\widehat{cZ}_{m-1}U_m''^{\dagger(1)}\mathbf{M}^{\dagger(1)})\ldots\\ &(\mathbf{R}^{\dagger(v+1)}U_1''^{\dagger(v+1)}\mathbf{F}_1^{\dagger(v+1)}\widehat{cZ}_1\ldots U_{m-1}''^{\dagger(v+1)}\mathbf{F}^{\dagger(v+1)}_{m-1}\widehat{cZ}_{m-1}U_m''^{\dagger(v+1)}\mathbf{M}^{\dagger(v+1)})],\end{aligned} \quad \text{(A16)}$$

where $\rho_{\text{in}} = \otimes_{k=1}^{v+1}\otimes_{i=1}^n |+\rangle_i^{(k)}\langle+|$, $\mathbf{R}^{(k)}$, $\mathbf{F}_j^{(k)}$, $\mathbf{M}^{(k)}$ are unitary matrices that act only on the qubits in the kth circuit and on the environment (which is the same for all the circuits) and $U_j''^{(k)} = \otimes_i U_{i,j}''^{(k)}$. Starting from here and using the same arguments as above, one can finally obtain equation (A1). □

Appendix B. Statement and proof of lemma 2

Lemma 1 shows that in our accreditation protocol the noise of the form N1 can be reduced to classically correlated collections of Pauli errors affecting the circuits. In this appendix we prove lemma 2, which states that all collections of Pauli errors can be detected with probability larger than 1/4. More formally:

Lemma 2. *For any collection of Pauli errors affecting a trap circuit, summing over all possible single-qubit gates in the trap circuit (i.e. over all possible sets $\{U_{i,j}\}$ output by Routine 2), the probability that the trap circuit outputs $\bar{s} = \bar{0}$ is at most 3/4.*

Proof. For a given collection of Pauli errors $\{\mathcal{P}_j\}_{j=0}^m$ affecting a trap circuit, the state of the trap circuit after the measurements is of the form

$$\rho^{\text{trap}}_{\text{out}}(\{\mathcal{P}_j\}) = \frac{1}{M_1\times\cdots\times M_{m-1}}\sum_{\mathcal{U}_1,\cdots,\mathcal{U}_m}\mathcal{P}_m\mathcal{U}_m\circ(\circ_{j=1}^{m-1}\mathcal{CZ}_j\mathcal{P}_j\mathcal{U}_j)\circ\mathcal{P}_0(\rho^{\text{trap}}_{\text{in}}), \quad \text{(B1)}$$

where $\rho^{\text{trap}}_{\text{in}} = \otimes_{i=1}^n|+\rangle_i\langle+|$, $\mathcal{CZ}_j$ is the entangling operation in band j, $\mathcal{P}_0, \mathcal{P}_m \in \{\mathcal{I}, \mathcal{Z}\}^{\otimes n}$, $\mathcal{P}_j \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes n}$ for all $j = 1,\ldots,m$ and M_j is the number of choices of $\mathcal{U}_j$. Note that each number M_j depends on the number of qubits connected by a cZ in band j of the trap circuit, see Routine 2.

In a trap circuit the gate $\mathcal{U}_1$ in the first band is of the form $\mathcal{U}_1 = \mathcal{V}_1\mathcal{H}^t$, where $\mathcal{V}_1$ implements a gate from $\{H, S\}^{\otimes n}$ (see step 2.1 of Routine 2) and $\mathcal{H}^t$ is the round of Hadamard gates activated at random (see step 3 of Routine 2). Similarly, for all $j = 2,\ldots,m-1$, $\mathcal{U}_j$ implements a gate belonging to the set $\{I, HS^\dagger, SH\}^{\otimes n}$. These gates undo the gates in previous band and implement new ones (see step 2.1 Routine 2 and figure 3), thus we can write them as $\mathcal{U}_j = \mathcal{V}_j\mathcal{V}_{j-1}^{-1}$—with each $\mathcal{V}_j$ implementing a gate from the set $\{H, S\}^{\otimes n}$. Finally, the gate $\mathcal{U}_m$ in the last band is of the form $\mathcal{U}_m = \mathcal{H}^t\mathcal{V}_{m-1}^{-1}$, where $\mathcal{V}_{m-1}^{-1}$ implements a gate from $\{H, S^\dagger\}^{\otimes n}$ and undoes the gate in band $m-1$ (see step 2.2 of Routine 2). Using this, we obtain

$$\rho^{\text{trap}}_{\text{out}}(\{\mathcal{P}_j\}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1\times\cdots\times N_{m-1}}\sum_{\mathcal{V}_1,\cdots,\mathcal{V}_{m-1}}\mathcal{P}_m\mathcal{H}^t\circ(\circ_{j=1}^{m-1}\mathcal{V}_j^{-1}\mathcal{CZ}_j\mathcal{P}_j\mathcal{V}_j)\circ\mathcal{H}^t\mathcal{P}_0(\rho^{\text{trap}}_{\text{in}}), \quad \text{(B2)}$$

where N_j is the number of possible choices of $\mathcal{V}_j$.

Using that $\mathcal{V}_j^{-1}\mathcal{CZ}_j\mathcal{V}_j = \mathcal{CX}_j$ is a tensor product of cX gates, the above state can also be rewritten as

$$\rho_{\text{out}}^{\text{trap}}(\{\mathcal{P}_j\}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1\times\cdots\times N_{m-1}}\sum_{\mathcal{V}_1,\cdots,\mathcal{V}_{m-1}}\mathcal{P}_m\mathcal{H}^t\circ(\circ_{j=1}^{m-1}\mathcal{CX}_j\mathcal{V}_j^{-1}\mathcal{P}_j\mathcal{V}_j)\circ\mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}}). \tag{B3}$$

Notice that each $\mathcal{CX}_j$ carries an implicit dependency on $\mathcal{V}_j$ (the orientation of the cX gates depends on $\mathcal{V}_j$, see Figure 3).

The probability that the trap outputs $\bar{s} = \bar{0}$ is

$$\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) = \langle +|^{\otimes n}\rho_{\text{out}}^{\text{trap}}(\{\mathcal{P}_j\})|+\rangle^{\otimes n}. \tag{B4}$$

To upper-bound this probability by 3/4, we first consider 'single-band' collections of errors, namely collections $\{\mathcal{P}_j\}$ such that $\mathcal{P}_{j_0}\neq\mathcal{I}$ for some $j_0\in\{0,\ldots,m\}$ and $\mathcal{P}_j=\mathcal{I}$ for all other $j\neq j_0$. For these collections, we prove that the probability that the output of the trap is the correct one $\bar{s}=\bar{0}$ is smaller than 1/2:

$$\text{prob}(\bar{s} = \bar{0}|\text{single} - \text{band collection}) \leqslant \frac{1}{2}. \tag{B5}$$

We prove this in Statement 1.

Next, we consider 'two-band' collections of errors. We obtain

$$\text{prob}(\bar{s} = \bar{0}|\text{two} - \text{band collection}) \leqslant \frac{3}{4}. \tag{B6}$$

We prove this in Statement 2. To obtain this bound, we *move* the two errors towards each other (i.e. we commute them with all the gates in the middle) and subsequently *merge* them, rewriting them as a single Pauli operator. The resulting Pauli operator is the identity $\mathcal{I}$ with probability c, or is a different operator with probability $1-c$. In the former case, the errors have canceled out with each other, while in the latter they have reduced to a single-band error. Importantly, in Statement 2 we prove that $c\leqslant 1/2$. This yields

$$\begin{aligned}\text{prob}(\bar{s} = \bar{0}|\text{two} - \text{band collection}) &= (1-c)\text{prob}(\bar{s} = \bar{0}|\text{single} - \text{band collection}) + c\ \text{prob}(\bar{s} = \bar{0}|\text{no error})\\ &\leqslant\frac{1-c}{2}+c=\frac{1}{2}+\frac{c}{2},\end{aligned} \tag{B7}$$

where we used $\text{prob}(\bar{s} = \bar{0}|\text{no error}) = 1$ and $\text{prob}(\bar{s} = \bar{0}|\text{single} - \text{band collection}) \leqslant 1/2$. Maximizing over $c\in[0,\ 1/2]$, we find

$$\text{prob}(\bar{s} = \bar{0}|\text{two} - \text{band collection}) \leqslant \max_{0\leqslant c\leqslant\frac{1}{2}}\left(\frac{1}{2}+\frac{c}{2}\right)=\frac{3}{4}. \tag{B8}$$

Finally, we generalise to collections affecting more than two bands. For three-band collections, again we move two of these Pauli operators towards each other and merge them. Doing this, the three-band collection reduces to a single-band one with probability $c\leqslant 1/2$ or to a two-band one with probability $1-c$. Thus, using the above results, we have

$$\begin{aligned}\text{prob}(\bar{s} = \bar{0}|\text{three} - \text{band collection}) &= (1-c)\text{prob}(\bar{s} = \bar{0}|\text{two} - \text{band collection}) + c\text{prob}(\bar{s} = \bar{0}|\text{single} - \text{band collection})\\ &\leqslant\frac{3(1-c)}{4}+\frac{c}{2}\leqslant\max_{0\leqslant c\leqslant\frac{1}{2}}\left(\frac{3(1-c)}{4}+\frac{c}{2}\right)=\frac{3}{4}.\end{aligned} \tag{B9}$$

This argument can be iterated: at any fixed h, if $\text{prob}(\bar{s} = \bar{0}|(h-2) - \text{band collection}) \leqslant 3/4$ and $\text{prob}(\bar{s} = \bar{0}|(h-1) - \text{band collection}) \leqslant 3/4$, then it can be easily shown that $\text{prob}(\bar{s} = \bar{0}|h - \text{band collection}) \leqslant 3/4$. We now complete the proof by proving Statement 1 and Statement 2.

Statement 1. Single-band collections are defined as follows:

$$\mathcal{P}_j\neq\mathcal{I}\text{ for } j=j_0\in\{0,\cdots,m\},\ \mathcal{P}_j=\mathcal{I}\text{ for all } j\neq j_0. \tag{B10}$$

If $j_0=0$, using $cX|{++}\rangle = |{++}\rangle$ and $cX|00\rangle = |00\rangle$, we have

$$\begin{aligned}\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) &= \frac{1}{2}\sum_{t=0,1}\langle +|^{\otimes n}\mathcal{H}^t\circ(\circ_{j=1}^{m}\mathcal{CX}_j)\circ\mathcal{H}^t\mathcal{P}_0|+\rangle^{\otimes n}\\ &=\langle +|^{\otimes n}\mathcal{P}_0|+\rangle^{\otimes n}=0,\end{aligned} \tag{B11}$$

since $\mathcal{P}_0\neq\mathcal{I}\in\{\mathcal{I},\ \mathcal{Z}\}^{\otimes n}$, and the same happens for $\mathcal{P}_m\neq\mathcal{I}$.

If $1\leqslant j_0\leqslant m-1$, we have

$$\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) = \frac{1}{N_{j_0}}\sum_{\mathcal{V}_{j_0}}\frac{1}{2}\sum_{t=0,1}\langle +|^{\otimes n}\mathcal{H}^t\mathcal{V}_{j_0}^{-1}\mathcal{P}_{j_0}\mathcal{V}_{j_0}\mathcal{H}^t|+\rangle^{\otimes n}, \tag{B12}$$

where we used again that $cX|{++}\rangle = |{++}\rangle$ and $cX|00\rangle = |00\rangle$. Notice that $\langle +|^{\otimes n}\mathcal{P}|+\rangle^{\otimes n}=0$ for all Pauli operators P whose Pauli-Z component is non-trivial, therefore $\sum_t\langle +|^{\otimes n}\mathcal{H}^t\mathcal{P}\mathcal{H}^t|+\rangle^{\otimes n}/2\leqslant 1/2$ for all

$\mathcal{P} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes n}/\mathcal{I}$. This yields

$$\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) \leqslant \frac{1}{N_{j_0}} \sum_{\mathcal{V}_{j_0}} \frac{1}{2} \leqslant \frac{1}{2}, \tag{B13}$$

where we used that $\mathcal{V}_{j_0}^{-1}\mathcal{P}_{j_0}\mathcal{V}_{j_0}$ is a Pauli operator for any $\mathcal{V}_{j_0}$, as this $\mathcal{V}_{j_0}$ implements a gate from the set $\{H, S\}^{\otimes n}$.

Statement 2. Two-band collections are defined as follows:

$$\mathcal{P}_j \neq \mathcal{I} \text{ for } j = j_1, j_2 \in \{0,\cdots,m\} (\text{with } j_1 < j_2),\ \mathcal{P}_j = \mathcal{I} \text{ for all } j \neq j_1, j_2. \tag{B14}$$

We can distinguish four classes of two-band collections:

(1) Errors in state preparation and entangling gates, i.e. $j_1 = 0$ and $1 \leqslant j_2 \leqslant m - 1$.

(2) Errors in entangling gates and measurements, i.e. $1 \leqslant j_1 \leqslant m - 1$ and $j_2 = m$.

(3) Errors in two different entangling gates, i.e. $1 \leqslant j_1 < j_2 \leqslant m - 1$.

(4) Errors in state preparation and measurements, i.e. $j_1 = 0$ and $j_2 = m$.

Errors in class 1 yield $\bar{s} = \bar{0}$ with probability at most 3/4. To prove this, we start by rewriting this probability as

$$\begin{aligned}
\text{prob}(\bar{s} = \bar{0}|\{P_j\}) &= \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1 \times \cdots \times N_{m-1}} \\
&\times \sum_{\mathcal{V}_1,\cdots,\mathcal{V}_{m-1}} \langle +|^{\otimes n}\mathcal{H}^t \circ (\circ_{j=j_2}^{m-1}\mathcal{CX}_j)\circ\mathcal{V}_{j_2}^{-1}\mathcal{P}_{j_2}\mathcal{V}_{j_2}\circ(\circ_{j=1}^{j_2-1}\mathcal{CX}_j)\circ\mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes n} \\
&= \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1 \times \cdots \times N_{m-1}}\sum_{\mathcal{V}_1,\cdots,\mathcal{V}_{j_2}} \langle +|^{\otimes n}\mathcal{H}^t\mathcal{V}_{j_2}^{-1}\mathcal{P}_{j_2}\mathcal{V}_{j_2}\circ(\circ_{j=1}^{j_2-1}\mathcal{CX}_j)\circ\mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes n},
\end{aligned} \tag{B15}$$

where we used $cX|{+}{+}\rangle = |{+}{+}\rangle$ and $cX|00\rangle = |00\rangle$. We now start from the case $j_2 = 1$. We then note that (i) if $n = 1$ (single-qubit circuit), all $\mathcal{P}_1 \in \{\mathcal{X}, \mathcal{Y}, \mathcal{Z}\}$ and all $\mathcal{P}_0 = \mathcal{Z}$ lead to

$$\text{prob}(\bar{s} = \bar{0}|\{P_j\}, 1\text{ qubit}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1}\sum_{\mathcal{V}_1\in\{\mathcal{H},\mathcal{S}\}}\langle +|\mathcal{H}^t\mathcal{V}_1^{-1}\mathcal{P}_1\mathcal{V}_1\mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}})|+\rangle \leqslant \frac{3}{4}, \tag{B16}$$

and (ii) if $n = 2$ (two-qubit circuit) and in band 1 the two qubits are connected by cZ, all $\mathcal{P}_1 \neq \mathcal{I} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes 2}$ and all $\mathcal{P}_0 \neq \mathcal{I} \in \{\mathcal{I}, \mathcal{Z}\}^{\otimes 2}$ lead to

$$\text{prob}(\bar{s} = \bar{0}|\{P_j\}, 2\text{qubits}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1}\sum_{\mathcal{V}_1\in\{\mathcal{H}\otimes\mathcal{S},\mathcal{S}\otimes\mathcal{H}\}}\langle +|^{\otimes 2}\mathcal{H}^t\mathcal{V}_1^{-1}\mathcal{P}_1\mathcal{V}_1\mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes 2} \leqslant \frac{3}{4}. \tag{B17}$$

The above inequalities for $n = 1$ and $n = 2$ can be proven using that H maps $\{X, Y, Z\}$ into $\{Z, Y, X\}$ under conjugation and S maps $\{X, Y, Z\}$ into $\{Y, X, Z\}$ under conjugation (apart from unimportant global phases). Extension to more than two qubits is as follows: If $\text{prob}(\bar{s} = \bar{0}|\{P_j\}, n_0\text{qubits}) \leqslant 3/4$, then tensoring one more qubit yields

$$\begin{aligned}
&\text{prob}(\bar{s} = \bar{0}|\{P_j\}, n_0 + 1\text{ qubits}) \\
&= \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1}\sum_{\mathcal{V}_1}(\langle +|^{\otimes n_0}\mathcal{H}_{1,\ldots,n_0}^t\mathcal{V}_{1|1,\ldots,n_0}^{-1}\mathcal{P}_{1|1,\ldots,n_0}\mathcal{V}_{1|1,\ldots,n_0}\mathcal{H}_{1,\ldots,n_0}^t\mathcal{P}_{0|1,\ldots,n_0}(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes n_0} \\
&\times \langle +|\mathcal{H}_{n_0+1}^t\mathcal{V}_{1|n_0+1}^{-1}\mathcal{P}_{1|n_0+1}\mathcal{V}_{1|n_0+1}\mathcal{H}_{n_0+1}^t\mathcal{P}_{0|n_0+1}(\rho_{\text{in}}^{\text{trap}})|+\rangle),
\end{aligned} \tag{B18}$$

where $\mathcal{H}_{1,\ldots,n_0}^t$, $\mathcal{V}_{1|1,\ldots,n_0}$, $\mathcal{P}_{1|1,\ldots,n_0}$ and $\mathcal{P}_{0|1,\ldots,n_0}$ are the components of $\mathcal{H}^t$, $\mathcal{V}_1$, $\mathcal{P}_1$ and $\mathcal{P}_0$ acting on qubits $\{1,\ldots,n_0\}$ and $\mathcal{H}_{n_0+1}^t$, $\mathcal{V}_{1|n_0+1}$, $\mathcal{P}_{1|n_0+1}$ and $\mathcal{P}_{0|n_0+1}$ the components acting on qubit $n_0 + 1$. Using that if $A_h, B_h \geqslant 0\,\forall\, h$, then $\sum_h A_h B_h \leqslant \sum_h A_h \sum_h B_h$, we obtain

$$\begin{aligned}
&\text{prob}(\bar{s} = \bar{0}|\{P_j\}, n_0 + 1\text{qubits}) \\
&\leqslant \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1}\sum_{\mathcal{V}_1}(\langle +|^{\otimes n_0}\mathcal{H}_{1,\ldots,n_0}^t\mathcal{V}_{1|1,\ldots,n_0}^{-1}\mathcal{P}_{1|1,\ldots,n_0}\mathcal{V}_{1|1,\ldots,n_0}\mathcal{H}_{1,\ldots,n_0}^t\mathcal{P}_{0|1,\ldots,n_0}(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes n_0}) \\
&\times \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1}\sum_{\mathcal{V}_1}(\langle +|\mathcal{H}_{n_0+1}^t\mathcal{V}_{1|n_0+1}^{-1}\mathcal{P}_{1|n_0+1}\mathcal{V}_{1|n_0+1}\mathcal{H}_{n_0+1}^t\mathcal{P}_{0|n_0+1}(\rho_{\text{in}}^{\text{trap}})|+\rangle) \leqslant \frac{3}{4} \times \frac{3}{4} \leqslant \frac{3}{4}.
\end{aligned} \tag{B19}$$

Tensoring two qubits connected by cZ yields the same bound, and this concludes the proof by induction for $j_2 = 1$. If $j_2 \in \{1,\ldots,m - 1\}$ the proof is similar, but the Pauli operator $\mathcal{V}_{j_2}^{-1}\mathcal{P}_{j_2}\mathcal{V}_{j_2}$ must be commuted with $\mathcal{CX}_1,\ldots,\mathcal{CX}_{j_2-1}$ (where we remember that each $\mathcal{CX}_j$ depends on $\mathcal{V}_j$). At fixed $\mathcal{V}_1,\ldots,\mathcal{V}_{j_2-1}$ it can be shown (with the same arguments as used for $j_2 = 1$, i.e. considering first the cases $n = 1$ and $n = 2$ and then generalizing to

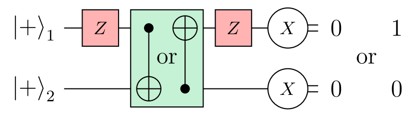

Figure 9. In this example, $\mathcal{P}_0 = \mathcal{P}_1 = Z_1$ (red gates) and $t = 0$. Due to identities B23, commuting $\mathcal{P}_1$ with the entangling gate (green box, cX gate with random orientation) make the two errors cancel out if qubit 1 is the control qubit. On the contrary, if qubit 1 is the target qubit, the errors do not cancel and cause a bit-flip of output s_1. Thus, for $t = 0$ these errors are detected with probability $1/2$. The same can be proven for $t = 1$ using identities B24, as well as for all other errors $\mathcal{P}_0$, $\mathcal{P}_1$.

$n > 2$) that summations over $\mathcal{V}_{j_2}$ and t yield an upper-bound of $3/4$. The upper-bound on $\text{prob}(\bar{s} = \bar{0}|\{P_j\}, n_0 + 1 \text{ qubits})$ follows by summing over all possible values of $\mathcal{V}_1,\ldots,\mathcal{V}_{j_2-1}$.

Errors in class 2 yield $\bar{s} = \bar{0}$ with probability at most $3/4$. This can be proven with the same arguments as for errors in class 1.

Errors in class 3 yield $\bar{s} = \bar{0}$ with probability at most $3/4$. To see this, consider first the case where the errors affect neighboring bands ($j_2 = j_1 + 1$), which yields

$$\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_{j_1}N_{j_1+1}}\sum_{\mathcal{V}_{j_1},\mathcal{V}_{j_1+1}}\langle +|^{\otimes n}\mathcal{H}^t\mathcal{V}_{j_1+1}^{-1}\mathcal{P}_{j_1+1}\mathcal{V}_{j_1+1}\mathcal{V}_{j_1}^{-1}\mathcal{P}_{j_1}\mathcal{V}_{j_1}\mathcal{H}^t|+\rangle^{\otimes n} \leqslant \frac{3}{4}. \tag{B20}$$

As for errors in class 1, this can be shown by proving that the bound holds for the single-qubit case and the two-qubit one, and then using induction. If the errors affect two non-neighboring bands ($j_2 \neq j_1 + 1$), we have

$$\text{prob}(\bar{s} = \bar{0}|\{\mathcal{P}_j\}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_{j_1}N_{j_2}}\sum_{\mathcal{V}_{j_1},\mathcal{V}_{j_2}}\langle +|^{\otimes n}\mathcal{H}^t\mathcal{V}_{j_2}^{-1}\mathcal{P}_{j_2}\mathcal{V}_{j_2}(\circ_{j=j_1+1}^{j_2-1}\mathcal{C}\mathcal{X}_j)\mathcal{V}_{j_1}^{-1}\mathcal{P}_{j_1}\mathcal{V}_{j_1}\mathcal{H}^t|+\rangle^{\otimes n} \leqslant \frac{3}{4}. \tag{B21}$$

To prove the inequality, one can commute $\mathcal{V}_{j_1}^{-1}\mathcal{P}_{j_1}\mathcal{V}_{j_1}$ (which is a Pauli operator) with the entangling operation and use the same arguments as for $j_2 = j_1 + 1$.

Finally, errors in class 4 yield

$$\text{prob}(\bar{s} = \bar{0}|\{P_j\}) = \frac{1}{2}\sum_{t=0,1}\frac{1}{N_1 \times \cdots \times N_{m-1}}\sum_{\mathcal{V}_1,\cdots,\mathcal{V}_{m-1}}\langle +|^{\otimes n}\mathcal{P}_m\mathcal{H}^t \circ (\circ_{j=1}^{m-1}\mathcal{C}\mathcal{X}_j) \circ \mathcal{H}^t\mathcal{P}_0(\rho_{\text{in}}^{\text{trap}})|+\rangle^{\otimes n} \leqslant \frac{1}{2}. \tag{B22}$$

To see this, consider first the case $t = 0$, and consider commuting $\mathcal{P}_0 \in \{\mathcal{I}, \mathcal{Z}\}^{\otimes n}/\mathcal{I}$ with all the gates in the circuit. Since all these gates are cX gates with random orientation, the identities

$$\begin{aligned} cX(Z_1 \otimes I_2) &= (Z_1 \otimes I_2)cX \\ cX(I_1 \otimes Z_2) &= (Z_1 \otimes Z_2)cX \\ cX(Z_1 \otimes Z_2) &= (I_1 \otimes Z_2)cX \end{aligned} \tag{B23}$$

ensure that every time time that a Pauli-Z error is commuted with a cX, this error becomes another error, chosen at random from two possible ones—figure 9. This can be used to prove that if $t = 0$, errors in class 4 are detected with probability larger than $1/2$. The same considerations apply to the case $t = 1$, where the identities

$$\begin{aligned} cX(X_1 \otimes I_2) &= (X_1 \otimes X_2)cX \\ cX(I_1 \otimes X_2) &= (I_1 \otimes X_2)cX \\ cX(X_1 \otimes X_2) &= (X_1 \otimes I_2)cX \end{aligned} \tag{B24}$$

must be used instead of identities B23. □

Appendix C. Proof of theorem 1

We start by using lemma 1 (appendix 1), which allows to reduce noise of the type N1 to classically correlated single-qubit Pauli errors and to write the joint state of target and trap circuits after all circuits have been implemented (for a fixed choice of single-qubit gates $\mathcal{U}_j^{(k)}$) as

$$\rho_{\text{out}}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}) = \sum_{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}}\sum_{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}}\frac{\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}}$$
$$\times \bigotimes_{k=1}^{v+1}\langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{C}\mathcal{Z}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)}) \circ \mathcal{P}_0^{(k)}(\rho_{\text{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}), \tag{C1}$$

where $\rho_{\text{in}} = \otimes_i|+\rangle_i\langle +|$, $\bar{s}^{(k)} = (s_1^{(k)},\ldots,s_n^{(k)})$ is the output of the kth circuit, $\mathcal{Z}^{\bar{s}^{(k)}}(\rho) = \otimes_i Z_i^{s_i^{(k)}}\rho Z_i^{s_i^{(k)}}$, and $\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})$ is the joint probability of a collection of Pauli errors $\{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}\}$ affecting the system, with $\mathcal{P}_1^{(k)},\ldots,\mathcal{P}_{m-1}^{(k)} \in \{\mathcal{I}, \mathcal{X}, \mathcal{Y}, \mathcal{Z}\}^{\otimes n}$ and $\mathcal{P}_0^{(k)}, \mathcal{P}_m^{(k)} \in \{I, \mathcal{Z}\}^{\otimes n}$ for all k. Summing over all possible choices of single-qubit gates we thus obtain the state of target and traps at the end of the protocol:

$$\begin{aligned}\rho_{\text{out}} &= \sum_{\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}} \text{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})\ \ \rho_{\text{out}}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})\\ &= \sum_{\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}} \text{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}) \sum_{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}}\ \sum_{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}} \frac{\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}}\\ &\quad\times \bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{CZ}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\text{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}),\end{aligned} \tag{C2}$$

where $\text{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})$ is the probability of single-qubit gates $\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}$ being chosen. Crucially, notice that the probability associated to each collection of Pauli errors does not depend on the specific choice of single-qubit gates $\mathcal{U}_j^{(k)}$. We can thus rewrite the above state as

$$\begin{aligned}\rho_{\text{out}} &= \sum_{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}}\ \sum_{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}} \frac{\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}} \sum_{\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}} \text{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})\\ &\quad\times \bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{CZ}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\text{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}).\end{aligned} \tag{C3}$$

Consider now the state

$$\begin{aligned}\rho_{\text{out}}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}) &= \sum_{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}} \frac{1}{2^{n(v+1)}} \sum_{\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}} \text{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})\\ &\quad\times \bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{CZ}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\text{in}})]|+\rangle^{\otimes n}(\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}})\end{aligned} \tag{C4}$$

corresponding to a fixed collection of Pauli errors $\{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}\}$ and assume that the Pauli errors only affect $\hat{v} \in \{1,\ldots,v+1\}$ circuits. In this case, as the Pauli errors do not depend on the single-qubit gates, they do not depend on the random number v_0 (which labels the position of the target circuit) nor on the random parameters in the trap circuits. Therefore, summing over all choices of single-qubit gates, the probability that the target is among the $\hat{v}$ circuits affected by noise is $\hat{v}/(v+1)$.

Next using lemma 2 (which states that summed over all possible choices of single-qubit gates, trap circuits output $\bar{s} = \bar{0}$ with probability at most 3/4) we have that if $\hat{v} - 1$ trap circuits are affected by noise, they all output $\bar{s} = \bar{0}$ with probability at most $(3/4)^{\hat{v}-1}$. We thus obtain

$$\rho_{\text{out}}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}) = \tilde{b}\ \tilde{\tau}_{\text{out}}^{\prime\text{tar}} \otimes \sigma_{\bar{s}=\bar{0}}^{\text{trap}} + (1-\tilde{b})(\tilde{l}\ \sigma_{\text{out}}^{\text{tar}} \otimes \sigma_{\bar{s}=\bar{0}}^{\text{trap}} + (1-\tilde{l})\ \tilde{\tau}_{\text{out}}^{\text{tar}} \otimes \sigma_{\bar{s}\neq\bar{0}}^{\text{trap}}), \tag{C5}$$

where $\tilde{\tau}_{\text{out}}^{\prime\text{tar}}$ ($\sigma_{\text{out}}^{\text{tar}}$) is the state of the target circuit if the target computation is (is not) among the $\hat{v}$ traps affected by noise, $\tilde{\tau}_{\text{out}}^{\text{tar}}$ is an arbitrary state for the target, $\sigma_{\bar{s}=\bar{0}}^{\text{trap}}$ is the state of the traps when they all output $\bar{s} = \bar{0}$, $\sigma_{\bar{s}\neq\bar{0}}^{\text{trap}}$ is an arbitrary state for the traps orthogonal to $\sigma_{\bar{s}=\bar{0}}^{\text{trap}}$, $0 \leqslant \tilde{l} \leqslant 1$ and

$$\tilde{b} \leqslant \frac{\hat{v}}{v+1}\left(\frac{3}{4}\right)^{\hat{v}-1}. \tag{C6}$$

For all $v \geqslant 3$ the RHS of the above upper-bound on $\tilde{b}$ is maximized by $\hat{v} = 3$, which yields

$$\tilde{b} \leqslant \max_{\hat{v}} \frac{\hat{v}}{v+1}\left(\frac{3}{4}\right)^{\hat{v}-1} = \frac{3}{v+1}\left(\frac{3}{4}\right)^{2} = \frac{\kappa}{v+1},\ \ \text{with}\,\kappa \approx \frac{1.7}{v+1}. \tag{C7}$$

As this is true for all collections of Pauli errors $\{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}\}$, we can sum over all possible collections (weighted by their respective probabilities) and obtain

$$\begin{aligned}\rho_{\text{out}} &= \sum_{\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}} \text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})\ \ \rho_{\text{out}}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})\\ &= b\ \tau_{\text{out}}^{\prime\text{tar}} \otimes \sigma_{\bar{s}=\bar{0}}^{\text{trap}} + (1-b)(l\ \sigma_{\text{out}}^{\text{tar}} \otimes \sigma_{\bar{s}=\bar{0}}^{\text{trap}} + (1-l)\ \tau_{\text{out}}^{\text{tar}} \otimes \sigma_{\bar{s}\neq\bar{0}}^{\text{trap}})\end{aligned} \tag{C8}$$

with $b \leqslant \kappa/(v+1)$ and $0 \leqslant l \leqslant 1$. The theorem is finally proven by replacing the state of the traps with that of the flag bit, namely by replacing $\sigma_{\bar{s}=\bar{0}}^{\text{trap}}$ with $|\text{acc}\rangle$ and $\sigma_{\bar{s}\neq\bar{0}}^{\text{trap}}$ with $|\text{rej}\rangle$.

Appendix D. Mesothetic verification protocol

D.1. Background

We will now define the notions of protocol and verifiability in the cryptographic setting. We start by defining quantum states as states belonging to the Hilbert space $\mathbb{H}_{ABC} = \mathbb{H}_A \otimes \mathbb{H}_B \otimes \mathbb{H}_C$, where $\mathbb{H}_A$ and $\mathbb{H}_B$ are Alice's and Bob's registers and $\mathbb{H}_C$ is a common register used to exchange qubits. The definition of protocol is as follows:

Definition 3 (Protocol). We define a q-step protocol on input $\rho_{\text{in}} \in \mathcal{L}(\mathbb{H}_{ABC})$ as a series of CPTP maps $\{\mathcal{E}_{ABC}^{(p)}\}_{p=1}^{q} = \{\mathcal{E}_{AC}^{(p)} \circ \mathcal{E}_{BC}^{(p)}\}_{p=1}^{q}$ yielding the state $\rho_{\text{out}} = \circ_{p=1}^{q} \mathcal{E}_{ABC}^{(p)}(\rho_{\text{in}})$.

In our protocol Alice verifies the correct implementation of the target circuit through the trap circuits. If Bob is honest, all these circuits output a default output and the flag bit is set to $|\text{acc}\rangle \in \{|0\rangle, |1\rangle\}$, otherwise it is likely that they return an incorrect output and the flag bit is set to $|\text{rej}\rangle = |\text{acc} \oplus 1\rangle$. We thus define verifiability as follows:

Definition 4 (Verifiability). Consider a protocol $\{\mathcal{E}_{ABC}^{(p)}\}_{p=1}^{q}$ on input ρ_{in}. The protocol is 'δ_{cr}-complete' if

$$D(\text{Tr}_{BC}[\circ_{p=1}^{q} \mathcal{E}_{ABC}^{(p)}(\rho_{\text{in}})].\ \sigma_{\text{out}}^{\text{tar}} \otimes |\text{acc}\rangle\langle\text{acc}|) \leqslant 1 - \delta_{\text{cr}}\ , \tag{D1}$$

where $\sigma_{\text{out}}^{\text{tar}}$ is the correct state of the target at the end of the protocol, $|\text{acc}\rangle$ is the state of the state of the traps indicating acceptance and $0 \leqslant \delta_{\text{cr}} \leqslant 1$ is called 'completeness'.

The protocol is 'ε_{cr}-sound' if for any set of CPTP maps $\{\widetilde{\mathcal{E}}_{BC}^{(p)}\}_{p=1}^{q}$ acting on Bob's register and on the channel, the state in Alice's register at the end of the protocol is

$$D(\text{Tr}_{BC}[\circ_{p=1}^{q}(\mathcal{E}_{ABC}^{(p)} \circ \widetilde{\mathcal{E}}_{BC}^{(p)})(\rho_{\text{in}})],\ l\ \sigma_{\text{out}}^{\text{tar}} \otimes |\text{acc}\rangle\langle\text{acc}| + (1-l)\tau_{\text{out}}^{\text{tar}} \otimes |\text{rej}\rangle\langle\text{rej}|) \leqslant \varepsilon_{\text{cr}}, \tag{D2}$$

where $0 \leqslant \varepsilon_{\text{cr}} \leqslant 1$ is called 'soundness', $0 \leqslant l \leqslant 1$, $\tau_{\text{out}}^{\text{tar}}$ is an arbitrary state and $|\text{rej}\rangle$ is orthogonal to $|\text{acc}\rangle$.

We thus model Bob's deviations as arbitrary CPTP maps acting on the system being processed and on Bob's private register.

We can now explain how to translate our protocol in the cryptographic setting.

D.2. Mesothetic protocol

We assume that Alice, the verifier, and Bob, the prover, want to implement interactively a circuit of the type of figure 2, with n qubits, depth m and single-qubit gates $\{U_{i,j}\}$. We start by defining the resources required by Alice:

- Alice's resources: a device that can receive n qubits from Bob, apply a single-qubit gate to each of them and send the qubits back to Bob. Alice's device must be able to execute all single-qubit gates contained in the target circuit, together with $H, S, S^\dagger$ (used in trap circuits) and X, Y, Z (used for the QOTP). We assume that the overall number of these gates is G.

For the protocol to always accept, Bob must possess a quantum computer that can prepare qubits in the $|+\rangle$ state, execute cZ gates, measure in the Pauli-X basis and store qubits.

We can now explain our protocol, which is formally presented in Box 4 in the Methods. Our protocol takes as input a description of the target circuit and the number v of trap circuits, both inputs also being known to Bob. Steps 1, 2 and 3 are classical and solely involve Alice's register, which at the end of step 3 (for a fixed choice of single-qubit gates $\mathcal{U}_1''^{(1)},\ldots,\mathcal{U}_m''^{(v+1)}$) is in the state

$$\rho_{A,\text{in}}(\mathcal{U}_1''^{(1)},\ldots,\mathcal{U}_m''^{(v+1)}) = [\otimes_{k=1}^{v+1} \otimes_{j=1}^{m} \otimes_{i=1}^{n} |U_{i,j}''^{(k)}\rangle\langle U_{i,j}''^{(k)}|] \otimes [\otimes_{k=1}^{v+1} \otimes_{i=1}^{n} |\alpha_{i,m}^{(k)}\rangle\langle\alpha_{i,m}^{(k)}|]. \tag{D3}$$

Here, $|U_{i,j}''^{(k)}\rangle$ is a classical description of the gate $U_{i,j}''^{(k)}$. Since Alice's device can execute G different gates, this classical description requires at most $\log_2(G)$ bits. Moreover, $|\alpha_{i,m}^{(k)}\rangle$ is a classical description of the bit $\alpha_{i,m}^{(k)}$.

In step 4 Alice and Bob interact as follows. For all circuits $k = 1,\ldots,v+1$, in step 4.1 Bob creates n qubits in the state $|+\rangle$. Then, for each band $j = 1,\ldots,m$, Bob sends all the qubits to Alice, Alice acts on each qubit i with $U_{i,j}^{(k)}$, Alice sends all the qubits back to Bob and Bob implements $\widehat{cZ}_j$. Finally, in step 4.3 Bob measures all the qubits and returns the outputs to Alice. The random but $\alpha_{i,m}^{(k)}$ coming from the QOPT in the last round of single-qubit gates bit-flips the outputs, therefore Bob transmits to Alice the bits $s_i^{(k)} \oplus \alpha_{i,m}^{(k)}$. Alice post-processes those bits and obtains $s_i^{(k)}$. At the end of step 4, the states in Alice and Bob's registers (if Bob is honest) are of the form

$$\begin{aligned} &\rho_{A,\text{out}}(\mathcal{U}_1''^{(1)},\ldots,\mathcal{U}_m''^{(v+1)}) = [\otimes_{k=1}^{v+1} \otimes_{j=1}^{m} \otimes_{i=1}^{n} |U_{i,j}''^{(k)}\rangle\langle U_{i,j}''^{(k)}|] \otimes [\otimes_{k=1}^{v+1} \otimes_{i=1}^{n} |s_i^{(k)}\rangle\langle s_i^{(k)}|] \\ &\rho_{B,\text{out}} = \otimes_{k=1}^{v+1} \otimes_{i=1}^{n}\ Z_i^{s_i^{(k)} \oplus \alpha_{i,m}^{(k)}} |+\rangle_i^{(k)} \langle +| Z_i^{s_i^{(k)} \oplus \alpha_{i,m}^{(k)}}. \end{aligned} \tag{D4}$$

The protocol ends with Alice checking whether the traps yield the correct output and setting the flag bit to $|\text{acc}\rangle$ or $|\text{rej}\rangle$.

We will now prove completeness and soundness for our mesothetic protocol (we refer the reader to the beginning of the appendix for notation). We begin with the assumption that Alice's device is noiseless. Soundness requires the following Lemmas:

Lemma D1. *Suppose that Alice's device is noiseless. For a fixed choice of single-qubit gates $\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}$ by Alice, tracing out Bob's private register and summing over all the random bits $\alpha_{i,j}^{(k)}$, $\alpha'^{(k)}_{i,j}$, $\gamma_i^{(k)}$ except $\alpha_{i,m}^{(k)}$ (see Routine 1), the state in Bob's register at the end of the protocol is of the form*

$$\rho_{B,\text{out}}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}) = \sum_{\substack{\bar{s}^{(1)},\ldots,\bar{s}^{(v+1)}\\ \mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}}} \frac{\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}}$$

$$\bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\bar{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{CZ}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\text{in}})]|+\rangle^{\otimes n} \times (\otimes_i Z_i^{s_i^{(k)}\oplus\alpha_{i,m}^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}\oplus\alpha_{i,m}^{(k)}}), \quad \text{(D5)}$$

where $\rho_{\text{in}} = \otimes_i|+\rangle_i\langle +|$, $\bar{s}^{(k)} = (s_1^{(k)},\ldots,s_n^{(k)})$ is the output of the kth *circuit, $\mathcal{Z}^{\bar{s}^{(k)}}(\rho) = \otimes_i Z_i^{s_i^{(k)}}\rho Z_i^{s_i^{(k)}}$ and* $\text{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})$ *is the joint probability of a collection of Pauli errors $\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}$ affecting the system, with $\mathcal{P}_1^{(k)},\ldots,\mathcal{P}_{m-1}^{(k)}\in\{\mathcal{I},\mathcal{X},\mathcal{Y},\mathcal{Z}\}^{\otimes n}$ and $\mathcal{P}_0^{(k)}$, $\mathcal{P}_m^{(k)}\in\{\mathcal{I},\mathcal{Z}\}^{\otimes n}$ for all* k.

Proof. (Sketch) We start proving the lemma for a protocol with a single circuit ($v = 1$), next we generalize to $v > 1$. For a fixed choice of gates $U''_{i,j}$ by Alice, the joint state of Alice's register A and Bob's register B before the final measurements is of the form

$$\rho_{AB}(\mathcal{U}''^{(1)}_1,\ldots,\mathcal{U}''^{(v+1)}_m) = [\otimes_{j=1}^m\otimes_{i=1}^n|U''_{i,j}\rangle\langle U''_{i,j}|]\otimes[\otimes_{k=1}^{v+1}\otimes_{i=1}^n|\alpha_{i,m}\rangle\langle\alpha_{i,m}|]\otimes \quad \text{(Alice's register)}$$
$$[\widetilde{\mathcal{E}}_m\mathcal{U}''_m\circ(\circ_{j=1}^{m-1}\widetilde{\mathcal{E}}_j\mathcal{U}''_j)\circ\mathcal{E}_0(\rho_{\text{in}})]. \quad \text{(Bob's register)}, \quad \text{(D6)}$$

where $\widetilde{\mathcal{E}}_j$ are Bob's deviations. Specifically, $\widetilde{\mathcal{E}}_0$ is Bob's deviation when he prepares ρ_{in} (step 4.1), for all $j = 1,\ldots,m-1$, $\widetilde{\mathcal{E}}_j$ is Bob's deviations when in step 4.2 he should execute $\widehat{cZ_j}$, $\widetilde{\mathcal{E}}_m$ is Bob's deviation before he measures the qubits (step 4.3). Without loss of generality, suppose now that Bob holds another register, which we call 'ancillary register' and denote with B_{anc}. Tensoring ρ_{in} with the state σ_{anc} in the register B_{anc}, for $j\in\{1,\ldots,m-1\}$ we can rewrite Bob's deviations $\widetilde{\mathcal{E}}_j$ as unitaries $\mathbf{F}_j\widehat{cZ_j}$, where $\mathbf{F}_j$ is a unitary matrix that acts on Bob's register B and on B_{anc}—for convenience we indicate unitary matrices acting both on Bob's register B and on B_{anc} in bold font. Similarly, we can replace $\widetilde{\mathcal{E}}_0$ with the unitary $\mathbf{R}$ and $\widetilde{\mathcal{E}}_m$ with the unitary $\mathbf{M}$. We thus obtain

$$\rho_{ABB_{\text{anc}}}(\mathcal{U}''^{(1)}_1,\ldots,\mathcal{U}''^{(v+1)}_m) = [\otimes_{j=1}^m\otimes_{i=1}^n|U''_{i,j}\rangle\langle U''_{i,j}|]\otimes[\otimes_{k=1}^{v+1}\otimes_{i=1}^n|\alpha_{i,m}\rangle\langle\alpha_{i,m}|]$$
$$[\otimes[\mathbf{M}\ U''_m\mathbf{F}_{m-1}\widehat{cZ}_{m-1}U''_{m-1}\ldots\mathbf{F}_1\widehat{cZ}_1U''_1\mathbf{R}(\rho_{\text{in}}\otimes\sigma_{\text{anc}})\mathbf{R}^\dagger U''^\dagger_1\widehat{cZ}_1\mathbf{F}_1^\dagger\ldots U''^\dagger_{m-1}\widehat{cZ}_{m-1}\mathbf{F}^\dagger_{m-1}U''^\dagger_m\mathbf{M}^\dagger]. \quad \text{(D7)}$$

Tracing out Alice's register and B_{anc} yields

$$\text{Tr}_{AB_{\text{anc}}}[\rho_{ABB_{\text{anc}}}(\mathcal{U}''^{(1)}_1,\ldots,\mathcal{U}''^{(v+1)}_m)]$$
$$= \text{Tr}_{B_{\text{anc}}}[\mathbf{M}\ U''_m\mathbf{F}_{m-1}\widehat{cZ}_{m-1}U''_{m-1}\ldots\mathbf{F}_1\widehat{cZ}_1U''_1\mathbf{R}(\rho_{\text{in}}\otimes\sigma_{\text{anc}})\mathbf{R}^\dagger U''^\dagger_1\widehat{cZ}_1\mathbf{F}_1^\dagger\ldots U''^\dagger_{m-1}\widehat{cZ}_{m-1}\mathbf{F}^\dagger_{m-1}U''^\dagger_m\mathbf{M}^\dagger]. \quad \text{(D8)}$$

This state is equal to the state ρ in equation (A4), provided that we replace σ_{anc} with ρ_E and $\text{Tr}_{B_{\text{anc}}}$ with $\text{Tr Tr}_E[\cdot]$, hence the lemma can be proven repeating the same calculations. The reader can verify that the same applies to the case $v > 1$. □

Lemma D2. *For any collection of Pauli errors affecting a trap circuit, summing over all possible single-qubit gates in the trap circuit (i.e. over all possible sets $\{U_{i,j}\}$ output by Routine 2), the probability that the trap circuit outputs $\bar{s} = \bar{0}$ is at most 3/4.*

As Alice choses all the gates in a trap circuits with Routine 2, the proof is the same as that of lemma 2. Using these two lemmas we now compute the classical state in Alice's register after all the circuits have been implemented:

Lemma D3. *Suppose that Alice's device is noiseless. Then, for any number $v\geqslant 3$ of trap computations, the state in Alice's register at the end of the protocol is of the form (see definition 4 for notation)*

$$\rho_{\text{out}} = b\ \tau'^{\text{tar}}_{\text{out}}\otimes|\text{acc}\rangle\langle\text{acc}| + (1-b)(l\ \ \sigma^{\text{tar}}_{\text{out}}\otimes|\text{acc}\rangle\langle\text{acc}| + (1-l)\tau^{\text{tar}}_{\text{out}}\otimes|\text{rej}\rangle\langle\text{rej}|), \quad \text{(D9)}$$

where $\tau'^{\text{tar}}_{\text{out}}$ is an incorrect state for the target, $0\leqslant b\leqslant\kappa/(v+1)$ and $\kappa = 3(3/4)^2\approx 1.7$.

Proof. After Bob has implements all the circuits and communicates all the outputs to Alice, Alice holds the same classical state as Bob (equation (D5)). After Alice post-processes the outputs this state becomes (see Lemma D1 for notation)

$$\rho_{A,\mathrm{out}}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}) = \sum_{\substack{\tilde{s}^{(1)},\ldots,\tilde{s}^{(v+1)}\\ \mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}}} \frac{\mathrm{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}}$$

$$\bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\tilde{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{C}\mathcal{Z}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\mathrm{in}})]|+\rangle^{\otimes n} \times (\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}). \tag{D10}$$

Crucially, notice that the probability associated to each deviation $\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}$ does not depend on the choice of single-qubit gates made by Alice. Thus, summing over all choices of single-qubit gates, this state becomes

$$\rho_{A,\mathrm{out}} = \sum_{\substack{\tilde{s}^{(1)},\ldots,\tilde{s}^{(v+1)}\\ \mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)}}} \frac{\mathrm{prob}(\mathcal{P}_1^{(1)},\ldots,\mathcal{P}_m^{(v+1)})}{2^{n(v+1)}} \sum_{\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)}} \mathrm{prob}(\mathcal{U}_1^{(1)},\ldots,\mathcal{U}_m^{(v+1)})$$

$$\bigotimes_{k=1}^{v+1} \langle +|^{\otimes n}[\mathcal{Z}^{\tilde{s}^{(k)}}\mathcal{P}_m^{(k)}\mathcal{U}_m^{(k)}\circ_{j=1}^{m-1}(\mathcal{C}\mathcal{Z}_j\mathcal{P}_j^{(k)}\mathcal{U}_j^{(k)})\circ\mathcal{P}_0^{(k)}(\rho_{\mathrm{in}})]|+\rangle^{\otimes n} \times (\otimes_i Z_i^{s_i^{(k)}}|+\rangle_i\langle +|Z_i^{s_i^{(k)}}). \tag{D11}$$

The lemma can now be proven following the same steps as in the proof of theorem 1. □

Using these three lemmas we can now calculate completeness and soundness for the mesothetic protocol and obtain theorem D1:

Theorem D1. *Suppose that Alice's device is noiseless. Then, for any number $v \geqslant 3$ of trap circuits, the mesothetic protocol is verifiable with*

$$\delta_{\mathrm{cr}} = 1 \text{ and } \varepsilon_{\mathrm{cr}} = \frac{\kappa}{v+1}, \tag{D12}$$

where $\kappa = 3(3/4)^2 \approx 1.7$.

Proof. Completeness can be proven trivially showing that for all circuits, the random Pauli gates used for the QOTP cancel between them, and that the effects of the QOTP in the last band are recovered by the classical post-processing made by Alice. Therefore, all the circuits implement the desired computation, and a correctly implemented protocol yields the correct output for the target and the state of the traps indicating acceptance.

To prove soundness we need to compute the trace distance in equation (D2). Using lemma D3 this trace distance becomes

$$D(b\ \tau'^{\mathrm{tar}}_{\mathrm{out}} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}| + (1-b)[l\ \ \sigma^{\mathrm{tar}}_{\mathrm{out}} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}| + (1-l)\tau^{\mathrm{tar}}_{\mathrm{out}} \otimes |\mathrm{rej}\rangle\langle\mathrm{rej}|],\ l\ \sigma^{\mathrm{tar}}_{\mathrm{out}} \otimes |\mathrm{acc}\rangle\langle\mathrm{acc}| + (1-l)\tau^{\mathrm{tar}}_{\mathrm{out}} \otimes |\mathrm{rej}\rangle\langle\mathrm{rej}|) \leqslant b \tag{D13}$$

and since $b \leqslant \kappa/(v+1)$ by lemma D3, we obtain soundness $\varepsilon_{\mathrm{cr}} = \kappa/(v+1)$. □

We can also compute completeness and soundness in the case where Alice's device suffers bounded and potentially gate-dependent noise:

Theorem D2. *Suppose that Alice's device is affected by bounded noise, i.e. that for all circuits $k \in \{1,\ldots,v+1\}$ and bands $j \in \{1,\ldots,m\}$ she applies $\mathcal{E}_j^{(k)}U_j^{(k)}$, where $\mathcal{E}_j^{(k)} = (1-r_j^{(k)})\mathcal{I} + r_j^{(k)}\mathcal{E}'^{(k)}_j$ for some arbitrary CPTP-map $\mathcal{E}'^{(k)}_j$ and number $0 \leqslant r_j^{(k)} < 1$. Then, for any number $v \geqslant 3$ of trap computations, the mesothetic protocol is verifiable with*

$$\delta_{\mathrm{cr}} = 1 \text{ and } \varepsilon_{\mathrm{cr}} = g\frac{\kappa}{v+1} + 1 - g, \tag{D14}$$

where $\kappa = 3(3/4)^2 \approx 1.7$, $g = \prod_{j,k} 1 - r^{(k)}_{\max,j}$ and $r^{(k)}_{\max,j}$ is the maximum error rate of the round of gates in band j of circuit k, where this maximum is taken over all choices gates for this round.

Proof. The proof of completeness is the same as for theorem D1. To compute soundness, we denote as $\rho^{\star}_{\mathrm{out}}$ the state in Alice's register at the end of a protocol run when Alice's device is noisy, and as ρ_{out} the state in Alice's register at the end of a protocol run when Alice's device is noiseless (Lemma D3). Indicating as $r^{(k)}_{\max,j}$ the maximum error rate for gates in band j, we rewrite this noisy map as $\mathcal{E}_j^{(k)} = (1-r^{(k)}_{\max,j})I + r^{(k)}_{\max,j}\mathcal{Q}_j^{(k)}$ for some other CPTP map $\mathcal{Q}_j^{(k)}$. This allows to rewrite the classical state in Alice's register at the end of the protocol as $g\rho_{\mathrm{out}} + (1-g)\rho^{\star}_{\mathrm{out}}$ and to obtain the upper-bound the trace distance. □

We conclude this appendix by showing how our protocol can be made blind. Blindness is a property exhibited by many cryptographic protocols [25] defined as follows:

Definition 5 (Blindness). A protocol $\{\mathcal{E}^{(p)}_{ABC}\} = \{\mathcal{E}^{(p)}_{AB} \otimes \mathcal{E}^{(p)}_{BC}\}$ on input ρ_{in} is blind if for any set of maps $\{\widetilde{\mathcal{E}}^{(p)}_{BC}\}$ acting on Bob's register and on the channel, the state in Bob's register at the end of the protocol leaks at most a constant function of the input.

Typically blind cryptographic protocols leak an upper-bound on the size of the target circuit.

As introduced above, our mesothetic protocol is not blind. Indeed, Bob has access to non-trivial information about the target circuit, such as the position of cZ gates. This is not a problem for our scopes, as verifiability only relies on Bob's ignorance of the number v_0, which is kept secret by Alice with the QOTP. Nevertheless, blindness may be required for maintaining the privacy of the users in future scenarios of delegated computations, thus it is important to understand whether our protocol can be turned into a blind protocol.

To obtain blindness, Alice (endowed with an ideal device) must recompile the target circuit into a normal form, for instance by recompiling the target circuit into a circuit of the type illustrated in figure 6 (which inspired to Brickwork States [52]). Then, instead of giving as input to the mesothetic protocol a classical description of the target circuit, she only needs to input the desired size of the circuit in normal form (n qubits, depth m).

Implementing target and traps on a circuit in normal form makes Alice's instructions to Bob independent from the specific target circuit that Alice wishes to verify. This allows to prove blindness:

Theorem D3 (Blindness with Circuit in Normal Form). *Suppose that Alice can apply noiseless single-qubit gates. If Alice rewrites the target circuit in normal form, the mesothetic protocol (with input the desired size of the circuit in normal form) only leaks an upper-bound on the size of the target circuit.*

Proof. To prove blindness we notice that during the protocol run Bob cannot retrieve any information about the the target circuit. Indeed, Bob's tasks are the same for all circuits (prepare the same input state, execute the same entangling gates and measure in the same basis) and these tasks do not depend on the target circuit, since this target is implemented on a circuit in normal form. Moreover, the only type of information that Bob receives from Alice during the implementation of the circuits are the qubits at step 4.2.1, but the QOTP prevents Bob from retrieving useful information: at any $k = 1,\ldots,v+1$ and $j = 1,\ldots,m$, if Bob sends to Alice a state $\rho_j^{(k)}$, Alice returns to him the state

$$U_j''^{(k)}\rho_j^{(k)}U_j^{\dagger\prime\prime(k)} = \otimes_{i=1}^{n} Z^{\alpha_{i,j}^{(k)}} X^{\alpha_{i,j}'^{(k)}} [U_j^{(k)} P_{j-1}\rho_j^{(k)} P_{j-1} U_j^{\dagger(k)}] X^{\alpha_{i,j}'^{(k)}} Z^{\alpha_{i,j}^{(k)}}, \tag{D15}$$

where P_{j-1} is the Pauli operator that undoes previous QOTP. Summing over all possible $\alpha_{i,j}^{(k)}$ and $\alpha_{i,j}'^{(k)}$ yields the completely mixed state. □

References

[1] Google Bristlecone
[2] IBM Quantum Experience
[3] Rigetti quantum computing
[4] Aaronson S and Arkhipov A 2011 The computational complexity of linear optics *Proc. 43rd Annual ACM Symp on Theory of Computing* pp 333–42
[5] Bremner M, Montanaro A and Shepherd D 2016 Average-case complexity versus approximate simulation of commuting quantum computations *Phys. Rev. Lett.* **117** 080501
[6] Gao X, Wang S-T and Duan L-M 2017 Quantum supremacy for simulating a translation-invariant ising spin model *Phys. Rev. Lett.* **118** 040502
[7] Terhal B 2018 Quantum supremacy, here we come *Nat. Phys.* **14** 530–1
[8] Dalzell A, Harrow A, Koh D and La Placa R 2018 How many qubits are needed for quantum computational supremacy? arXiv:1805.05224
[9] Neill C *et al* 2017 A blueprint for demonstrating quantum supremacy with superconducting qubits *Science* **360** 195–9
[10] Preskill J 2018 Quantum computing in the NISQ era and beyond *Quantum* **2** 79
[11] Gottesman D 1999 *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics* ed S P Corney, R Delbourgo and P D Jarvis (Cambridge, MA: International Press) pp 32–43
[12] Bravyi S and Gosset D 2017 Improved classical simulation of quantum circuits dominated by Clifford gates *Phys. Rev. Lett.* **116** 250501
[13] Bennink R *et al* 2017 Unbiased simulation of near-Clifford quantum circuits *Phys. Rev.* A **95** 062337
[14] Boixo S *et al* 2018 Characterizing quantum supremacy in near-term devices *Nat. Phys.* **14** 595–600
[15] Villalonga B *et al* 2019 A flexible high-performance simulator for the verification and benchmarking of quantum circuits implemented on real hardware *npj Quantum Inf.* **5** 86
[16] Emerson J, Alicki R and Zyczkowski K 2005 Scalable noise estimation with random unitary operators *J. Opt. B: Quantum Semiclass. Opt.* **7** S347
[17] Knill E *et al* 2007 Randomized benchmarking of quantum gates *Phys. Rev.* A **77** 012307

[18] Dankert C, Cleve E, Emerson J and Livine E 2009 Exact and approximate unitary 2-designs: constructions and applications *Phys. Rev.* A **80** 012304
[19] Magesan E, Gambetta J and Emerson J 2011 Scalable and robust randomized benchmarking of quantum processes *Phys. Rev. Lett.* **106** 180504
[20] Erhard A *et al* 2019 Characterizing large-scale quantum computers via cycle benchmarking arXiv:1902.08543
[21] Combes J, Granade C, Ferrie C and Flammia S 2017 Logical randomized benchmarking arXiv:1702.03688
[22] Sanders Y, Wallman J and Sanders B 2016 Bounding quantum gate error rate based on reported average fidelity *New J. Phys.* **18** 012002
[23] Wallman J 2018 Randomized benchmarking with gate-dependent noise *Quantum* **2** 47
[24] Merkel S, Pritchett E and Fong B 2018 Randomized benchmarking as convolution: Fourier analysis of gate dependent errors arXiv:1804.05951
[25] Gheorghiu A, Kapourniotis T and Kashefi E 2018 Verification of quantum computation: An overview of existing approaches *Theory Comput. Syst.* **63** 715–808
[26] Aharonov D, Ben-Or M and Eban E 2010 Interactive proofs for quantum computations *Innovations in Computer Science—ICS 2010, Proc. (Tsinghua University, Beijing, China, 5–7 January 2010)* pp 453–69
[27] Fitzsimons J and Kashefi E 2017 Unconditionally verifiable blind computation *Phys. Rev.* A **96** 012303
[28] Barz S, Fitzsimons J, Kashefi E and Walther P 2013 Experimental verification of quantum computations *Nat. Phys.* **9** 727–31
[29] Kapourniotis T and Datta A 2019 Nonadaptive fault-tolerant verification of quantum supremacy with noise *Quantum* **3** 164
[30] Kashefi E and Wallden P 2017 Optimised resource construction for verifiable quantum computation *J. Phys. A: Math. Theor.* **50** 145306
[31] Ferracin S, Kapourniotis T and Datta A 2017 Reducing resources for verification of quantum computations *Phys. Rev.* A **98** 022323
[32] Aharonov D, Ben-Or M, Eban E and Mahadev U 2017 Interactive proofs for quantum computations arXiv:1704.04487
[33] Broadbent A 2018 How to verify a quantum computation *Theory Comput.* **14** 1–37
[34] Hayashi M and Morimae T 2015 Verifiable measurement-only blind quantum computing with stabilizer testing *Phys. Rev. Lett.* **115** 220502
[35] Morimae T and Fitzsimons J 2018 Post hoc verification with a single prover *Phys. Rev. Lett.* **120** 040501
[36] Hangleiter D, Kliesch M, Schwarz M and Eisert J 2017 Direct certification of a class of quantum simulations *Quantum Sci Technol.* **2** 015004
[37] Markham D and Krause A 2018 A simple protocol for certifying graph states and applications in quantum networks arXiv:1801.05057
[38] Takeuchi Y, Mantri A, Morimae T, Mizutani A and Fitzsimons J 2019 Resource-efficient verification of quantum computing using Serfling's bound *npj Quantum Inf.* **5** 27
[39] Mahadev U 2018 Classical verification of quantum computations *2018 IEEE 59th Annual Symp. on Foundations of Computer Science*
[40] Reichardt B W, Unger F and Vazirani U 2012 A classical leash for a quantum system: command of quantum systems via rigidity of CHSH games arXiv:1209.0448
[41] Gheorghiu A, Kashefi E and Wallden P 2015 Robustness and device independence of verifiable blind quantum computing *New J. Phys.* **17** 083040
[42] McKague M 2016 Interactive proofs for BQP via self-tested graph states *Theory Comput.* **12** 1–42
[43] Fitzsimons J and Hajdusek M 2015 Post hoc verification of quantum computation arXiv:1502.02563
[44] Natarajan A and Vidick T 2016 Robust self-testing of many-qubit states *Proc. of STOC '17* 2017 pp 1003–15
[45] Harty T P, Allcock D T C, Ballance C J, Guidoni L, Janacek H A, Linke N M, Stacey D N and Lucas D M 2014 High-fidelity preparation, gates, memory, and readout of a trapped-ion quantum bit *Phys. Rev. Lett.* **113** 220501
[46] Harty T P, Sepiol M A, Allcock D T C, Ballance C J, Tarlton E K and Lucas D M 2016 High-fidelity trapped-ion quantum logic using near-field microwaves *Phys. Rev. Lett.* **117** 140501
[47] Barends R *et al* 2014 Superconducting quantum circuits at the surface code threshold for fault tolerance *Nature* **508** 500–3
[48] Sheldon S, Magesan E, Chow J and Gambetta J 2016 Procedure for systematically tuning up cross-talk in the cross-resonance gate *Phys. Rev.* A **93** 060302(R)
[49] Aharonov D, Kitaev A and Nisan N 1999 Quantum circuits with mixed states *Proc. Thirtieth Annual ACM Symp. on Theory of Computation (STOC)'*
[50] Nielsen M and Chuang I 2000 *Quantum Computation and Quantum Information* 10th Anniversary edn (Cambridge: Cambridge University Press)
[51] Wallman J and Emerson J 2016 Noise tailoring for scalable quantum computation via randomized compiling *Phys. Rev.* A **94** 052325
[52] Broadbent A, Fitzsimons J and Kashefi E 2009 Universal blind quantum computation *Proc. 50th Annual IEEE Symp. on Foundations of Computer Science* pp 517–26
[53] Fitzsimons J 2017 Private quantum computation: an introduction to blind quantum computing and related protocols *npj Quantum Inf.* **3** 23
[54] Fujii K and Hayashi M 2017 Verifiable fault-tolerance in measurement-based quantum computation *Phys. Rev.* A **96** 030301